US Army Corps
of Engineers
Construction Engineering
Research Laboratory

USACERL Technical Manuscript N-89/18
September 1989

2

AD-A213 823

Proceedings of the 1988 Geographical Resource Analysis Support System (GRASS) User Group Meeting

Edited by:
Robert C. Lozar

The Geographical Resource Analysis Support System (GRASS) is a land management support tool originally developed to help military installations ensure realism in training while conserving the environment. Since its successful implementation in the military community, GRASS has seen wide-spread acceptance in both the Government and private sector.

This proceedings contains papers from the 1988 Anuual GRASS User Group Meeting which was held at USACERL in Champaign, IL. The papers represent a variety of interests. They have been grouped under three general topic areas: Applications, Data Concerns, and Integration of Grass With Other Packages.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704 0188
Exp Date Jun 30 1986

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b RESTRICTIVE MARKINGS:

2a SECURITY CLASSIFICATION AUTHORITY:

2b DECLASSIFICATION / DOWNGRADING SCHEDULE:

3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBER(S): USACERL Technical Manuscript N-89/18

5 MONITORING ORGANIZATION REPORT NUMBER(S):

6a NAME OF PERFORMING ORGANIZATION: U.S. Army Construction Engr Research Laboratory

6b OFFICE SYMBOL (If applicable): CECER-EN

6c ADDRESS (City, State, and ZIP Code): P.O. Box 4005, Champaign, IL 61824-4005

7a NAME OF MONITORING ORGANIZATION:

7b ADDRESS (City, State, and ZIP Code):

8a NAME OF FUNDING / SPONSORING ORGANIZATION: Ofc of the Secretary of Defense

8b OFFICE SYMBOL (If applicable):

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: Intra-Army Order 22 MSS-88-088, "Enhancement of the Early Environmental (Cont'd)

8c ADDRESS (City, State and ZIP Code):

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification): Proceedings of the 1988 Geographical Resource Analysis Support System (GRASS) User Group Meeting

12 PERSONAL AUTHOR(S): Edited by Lozar, Robert C.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
final	FROM ____ TO ____	1989, September	76

16 SUPPLEMENTARY NOTATION: Copies are available from the National Technical Information Services Springfield, VA 22161

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
08	06	

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Symposia; User Group

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The Geographical Resource Analysis Support System (GRASS) is a land management support tool originally developed to help military installations ensure realism in training while conserving the environment. Since its successful implementation in the military community, GRASS has seen widespread acceptance in both the Government and private sector.

This proceedings contains papers from the 1988 Anuual GRASS User Group Meeting which was held at USACERL in Champaign, IL. The papers represents a variety of interests. They have been grouped under three general topic areas: Applications, Data Concerns, and Integration of Grass With Other Packages.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL: Dana L. Finney

22b TELEPHONE (Include Area Code): (217) 352-6511 (x389)

22c OFFICE SYMBOL: CECER-IMT

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

Block 9 (Cont'd)

Warning System"

FOREWORD

This project was funded by the Office of the Secretary of Defense under Intra-Army Order 22 MSS-88-088, "Enhancement of the Early Environmental Warning System," dated September 1988. The proceedings were prepared by the U.S. Army Construction Engineering Research Laboratory Environmental Division (USACERL-EN). Dr. R.K. Jain is Chief of USACERL-EN.

COL Carl O. Magnell is Commander and Director of USACERL, and Dr. L.R. Shaffer is Technical Director.

Preface

The Geographical Resource Analysis Support System (GRASS) is a land management support tool originally developed to help military installations ensure realism in training while conserving the environment. Since its successful implementation in the military community, GRASS has seen widespread acceptance in both the Government and private sector.

The U.S. Army Construction Engineering Research Laboratory (USACERL) developed and tested GRASS, which is a geographic information system (GIS). Now, along with a formal GRASS Steering Committee and other service agencies, USACERL is providing support to users. Each year the Steering Committee sponsors a user group meeting for information exchange: other help is available through workshops, and online mail service called GRASSNET, and a periodical newsletter, GRASS-CLIPPINGS.

At the 1987 Annual User Group Meeting, response to the call for papers was so favorable and the quality of presentations so impressive that the organizers decided to publish the papers from future meetings. This proceedings contains papers from the 1988 Annual GRASS User Group Meeting which was held at USACERL in Champaign, IL.

The papers represent a variety of interests. They have been grouped under three general topic areas: Applications, Data Concerns, and Integration of Grass With Other Packages.

Papers in the Applications group illustrate the versatility of the GRASS software. Gary Waggoner of the National Park Service (NPS) outlines a procedure to define road corridors using GRASS. Two papers deal with sampling design: Susan Stitt, NPS, applies GRASS for determining forest locations to sample in evaluating the effects of air quality on vegetation; Steven Warren, USACERL uses GRASS to define areas where data collection sites should be distributed to ensure nonbiased data. GRASS applications in archeology are also reported. Ishmael Williams of the Arkansas Archeological Survey (AAS) describes how he uses GRASS to organize his data to reveal patterns of the Caddo Indian habitation sites. Jamie Lockhart, AAS, uses GRASS to coordinate and represent statistical and ordinal archeological information over an eight-state region. Pamela Thompson,

USACERL, demonstrates the versatility of soils information when handled with the GRASS tools as well as several potential new applications. David Hasting of the National Oceanic and Atmospheric Administration user GRASS to check the quality and consistency of the spatial data for which his agency is responsible.

Presentations in the Data Concerns group address the information base, which critical if a GIS like GRASS is to be useful. Since data collection is expensive, it is important to know which data layers to implement, and this is the topic of a paper by Robert Lozar of USACERL. Richard Franchek, U.S. Soil Conservation Service (SCS), describes and example data set for training personnel at SCS, where GRASS will soon be implemented in all county offices. Margaret Mayers of SPOT Image Corp. discusses the link between satellite data and GRASS--an important capability since satellite images can provide valuable data.

GRASS is designed to allow easy interface with other software packages for flexibility. Papers in the GRASS Integration group attest to the success users are having in this area. Sandra Parker, AAS, proposes linking a statistical package with GRASS and remotely sensed data to make the results more immediately understandable to professionals. Sanford Fidell and colleagues from BBN Systems describe how they ported sections of GRASS version 2.0 to an IBM-AT with MS-DOS and their integration with a data base manager (DBM) for evaluating aircraft noise. James Farley, AAS, discusses the interface of GRASS with a UNIX-based DBM called Informix. Finally, Ken Gardels, at the University of California-Berkeley, identifies the pros and cons of placing GRASS in a new standard graphics environment called X-Windows.

In the short time since its inception, the potential for GRASS has grown far beyond initial expectations. Each year the User Group Meeting unveils more and more benefits from using this program; the enthusiasm users have for GRASS is evident in these papers. Today GRASS faces and exciting future as it expands and sees adoption by a variety of new organizations. You are invited to share your experiences at the next User Group Meeting and to learn how others are using GRASS.

For more information on GRASS or the Annual User Group Meeting, contact the GRASS Support Center at USACERL, (217) 373-7220.

CONTENTS

CONTENTS (Cont'd)

APPLICATIONS

Analysis of Alternative Road Alignments using GRASS 3.0

Gary S. Waggoner

Geographic Information Systems Division
National Park Service
P.O. Box 25287
Denver, CO 80225-0287

ABSTRACT

A "real life" example using GRASS 3.0 GIS technology to aid planners in developing a new park entrance road alignment at Great Basin National Park, Nevada is presented. Analyses and plots have been field tested by National Park Service and Federal Highway Administration planners. GRASS 3.0 tools including distance, weight, Gcost, Gdrain, and Glos were employed in the analyses and are discussed in light of the field checking results.

Great Basin National Park was established on October 27, 1986 making it the newest national park in the National Park System. Located in White Pine County, in east-central Nevada, the park occupies 77,109 acres in the South Snake Range. It was established "to preserve for the benefit and inspiration of the people a representative segment of the Great Basin of the Western United States possessing outstanding resources and significant geological and scenic values." (Public Law 99-565). The National Park Service (NPS) is currently preparing a general management plan for the park (NPS, 1988) and a GIS data base including multiple resource themes encompassing a 928 square mile area is being created to support this activity.

One of the proposals being considered is the building of a Visitor Center in the northeastern portion of the park near Baker Creek. The NPS in cooperation with the Federal Highway Administration (FHA) is charged with developing a new park entrance road and requested the GIS Division's support in this undertaking.

Since the GIS Division was already conducting a beta test of the new GRASS 3.0 version, we decided to test several of the new capabilities with a "real life" experiment which would be field tested to give some practical evaluation and feedback.

The author had several meetings with NPS planners to fully discuss the intent and parameters involved in the test. After fully describing the hypothesis of the experiment, criteria for the selection of a road corridor and a geographical window were decided upon and GRASS 3.0 operations were begun. The planners indicated that three major criteria were important in these road alignment considerations:

1. Slopes should be 7 degrees or less
2. Stream crossings should be minimized and where necessary should be at right angles to the stream
3. The entrance road should be hidden as much as possible from the view of a visitor at the proposed new visitor center

Vegetation was not considered a significant criterion because the vegetation throughout the area of concern is a fairly homogeneous sagebrush-shadscale-grass association. Narrow riparian zones do occur but are automatically incorporated into the 30 meter cell size used to delineate the small streams in the area.

Using 1:250,000 scale digital topographic (DMA) data that had previously been processed into elevation, slope and aspect 90 meter cell data files, the slope data were reclassed into two classes, 0-7 degree slopes and greater than 7 degree slopes. Stream data were obtained from digital line graph (DLG) data at the 1:100,000 scale and rasterized into 30 meter cells.

Viewshed data were obtained by running Glos on the elevation data mentioned above. A UTM coordinate pair was digitized from a mylar 1:24,000 scale quadrangle on which the planners had located the proposed site of the Visitor Center. Glos was run to extend throughout the predetermined window of interest (9000 meters) as viewed from a point 50 feet above ground level. The height was determined to conservatively estimate the maximum viewshed from the highest potential point of development at the proposed visitor center.

Weight was next used to create a cell file which integrated all of the environmental criteria. The planners assisted in establishing relative weights for the environmental variables:

SLOPE:	0 - 7 degrees	=	0
	> 7 degrees		10
STREAMS:	non stream	=	0
	stream	=	10
VIEWSHED:	unseen area	=	0
	seen area	=	5

Weights were integrated in an additive fashion. Class values resulting from executing weight were 0, 5, 10, 15, 20, and 25 reflecting all the various combinations of variable weights. The resultant can be described as an "environmental cost map" depicting the interaction of all the variables considered in the analysis.

Gcost was then run on the "environmental cost map." Gcost evaluates a cost surface relative to a starting point. It establishes the starting point as a "low point" which the user selects by entering a UTM coordinate pair. This point serves as one of two or more points to be connected in the next step of the process ie. Gdrain. A cell map similar to an elevation map is created which evaluates the environmental cost of getting to each and every cell in the window from this starting point. The point was selected by the planners to represent an area near an existing road which would be used to access the proposed Visitor Center. Once again the point was accurately delineated on a mylar 1:24,000 scale quadrangle sheet and digitized to obtain accurate UTM coordinates.

Following the creation of the Gcost evaluation surface, Gdrain was employed to operate on this file. A second point was selected at the proposed intersection with the existing State Highway 48, approximately 2 miles south of Baker, Nevada. This intersection point was digitized from the mylar quadrangle sheet and entered as a required Gdrain variable in UTM coordinates.

Gdrain is essentially a gravity flow model and allows the user to connect designated points along the path of least resistance (or least environmental cost). It analyzes how an imaginary raindrop would drain from any beginning point to the low point on the map. If the "topography" of the map is in actuality a synthesis of environmentally sensitive factors, then the resulting drainage path is an environmental least cost path or corridor connecting the selected points.

The result of the Gdrain operation produced a corridor for acceptable road development comprised of sections of narrow road alignment, 30 meters wide, and sections of broader zones of acceptable road corridor hundreds of meters wide. In order to select a specific, 30 meter wide road alignment, within the broad zones of the acceptable road corridor, the author decided to use a buffering approach which would force the selection of the shortest route within the acceptable zone.

Distance was used to develop a buffer surface with 237 concentric, 30 meter wide, rings emanating from the initial digitized point ie. the "low point" from the Gcost surface. Each ring was weighted at its ring order number away from the center point eg. 1, 2, 3, ... Additionally, the unacceptable road alignment zone was weighted at 500 while the acceptable road corridor zone was weighted at 0. Weights were integrated

in an additive fashion.

The resultant environmental cost surface map was operated on in a similar fashion as described above with both Gcost and Gdrain. The result of these sequential operations was the creation of a single, 30 meter wide, 6.5 mile long alignment connecting the two points and occurring only within the acceptable road corridor zone previously determined.

The narrow road alignment was then added to the initial broader zone corridor using patch. This patched file was in turn patched into a cell file with existing roads and trails. A plot at 1:12,000 scale was then produced on mylar to overlay a topographic map which the NPS and FHA planners used in the field to verify the results of the analysis.

During the last week in September, NPS and FHA planners went to Great Basin National Park to evaluate the results of the GRASS 3.0 road alignment analysis. After extensive field work, checking numerous sites along the GRASS 3.0-generated, proposed road alignment, planners found the alternative to be acceptable and to meet the criteria used in the model. They were very encouraged by the results of the field test stating that "the corridor mapping was extremely accurate and very helpful." (Goodrich, 1988). Subsequently, the FHA planner in charge of developing the final road alignment specifications expressed his desire to work with our office combining the Great Basin data in GRASS 3.0 with the technical engineering data that the FHA CADD system produces. "The applications could be very beneficial to both agencies." (Goodrich 1988). Our office will be pursuing this opportunity

NPS planners have been further encouraged to use GIS technology by the success of this application. The entire analysis was accomplished within a few days and was comprehensive and thorough in its use of road alignment selection criteria. In spite of the relatively gross digital topographic data used in the analysis, highly useful, unbiased information has been generated in a timely fashion.

This application is also significant because it further emphasizes the usefulness of GIS technology in alternative formulation, in addition to environmental impact assessment, where it has been most frequently used in the NPS. Although the use of GIS technology is far from routine in the NPS, successful applications such as this one help to build a track record and develop realistic expectations in the eyes of park management. Once realistic expectations from management can be met by GIS technology advancement, routine use of GIS will occur. Based on the success and acceptance of this application, GRASS 3.0 has moved the National Park Service closer to routine use of GIS.

REFERENCES

Great Basin National Park Act of 1986. (Public Law 99-565-- October 27,1986).

National Park Service. U. S. Department of the Interior. 1988. Great Basin National Park/Nevada, Alternatives Workbook, Background and Alternatives. NPS D-12. July 1988.

Goodrich, Terry. 1988. Memorandum To: Phil Wondra, WASO, Geographic Information Systems Division; Reference: Great Basin National Park, Road System Evaluation; Subject: Request for Technical Assistance. October 17, 1988. Western Team, Denver Service Center, National Park Service.

SELECTION OF LAND CONDITION INVENTORY SITES USING THE GEOGRAPHIC RESOURCES ANALYSIS SUPPORT SYSTEM

Steven D. Warren, Mark O. Johnson, Victor E. Diersing and William D. Goran

US Army Construction Engineering Research Laboratory
Environmental Division
P.O. Box 4005
Champaign, Illinois 61820

ABSTRACT

Selection of representative sites for land condition inventory can be a time-consuming and subjective task. A procedure is currently being developed by the US Army Construction Engineering Research Laboratory to remove subjectivity and automate the site selection process for military installations. The procedure incorporates satellite imagery and digital soil surveys in the Geographic Resources Analysis Support System environment. A military installation is stratified by imagery-derived landcover categories and soil series. Inventory sites are allocated in a stratified random fashion based on the percentage of the installation occupied by the various unique landcover/soil series categories.

Background

The US Army manages approximately 4.5 million hectares of forest and rangeland in the United States. Increasing demands for more frequent and larger scale military training exercises compatible with modern weaponry systems have taken a toll on the land and natural resources. Some military installations have experienced undesirable shifts in plant species composition, reductions in vegetative cover and accelerated soil erosion. As a result, field training realism has diminished and the longevity of the land for military training purposes has been threatened.

In an effort to halt the degradation of military land, the US Army Construction Engineering Research Laboratory is developing an Integrated Training Area Management (ITAM) program (1). The program seeks to enhance natural resource conservation and realistic field training through the integration of military training requirements with environmental awareness education, land rehabilitation efforts and land-use planning based on the capacity of the land to support various forms of military training. Effective land management is dependent, in large part, on accurate assessment of the quantity and quality of available resources. Therefore, a major thrust of the ITAM program is to inventory the current condition of Army training lands in terms of factors such as soil erosion and concealment resources, and monitor trends in land condition over time through a standardized procedure known as Land Condition-Trend Analysis (LCTA) (2).

LCTA incorporates on-the-ground sampling of soils, topography and vegetation. Vegetation is evaluated with both point-intercept and belt transect methods and requires a minimum area of 100m x 6m. Selection of representative sampling sites for land condition inventory can be a time-consuming and highly subjective task. The purpose of this research is to develop an automated, objective procedure for selec-

tion of land condition inventory sites. The procedure utilizes satellite imagery, soil surveys and the Geographic Resources Analysis Support System (GRASS).

Discussion

The first step in the site selection process is to acquire a SPOT satellite image of a given military installation. Ideally, the image should be taken at the time of year when perennial plant growth is at a peak. Based on reflectance in the green, red and near infra-red spectral wavelength bands, an unsupervised classification is performed. All land areas outside of the installation boundary are masked from the satellite image prior to the classification to prevent influence from extraneous land cover types. A maximum of 20 spectrally unique landcover categories are permitted. Due to the nature of remotely sensed imagery, the landcover categories are sensitive primarily to the amount of vegetative cover and gross physiognomic differences in plant communities and, to a lesser degree, plant species composition.

The resulting landcover data layer provides an initial stratification for the site selection process. However, a single spectrally-recognized landcover category may actually represent more than one distinct plant community. This may be due to the limitation of 20 landcover categories imposed on the unsupervised classification algorithm or may result from the occurrence of more than one plant community with very similar spectral reflectance characteristics. In the latter case, differences in plant species composition are often correlated to differences in the soils that support the vegetation. Therefore, a secondary stratification based on soil series is appropriate.

Within GRASS a digital soil series data layer is superimposed on the landcover data layer. A GRASS "macro" algorithm has been written that causes the computer to recognize each unique landcover/soil combination as a separate category. The unique combination of a landcover category with a soil series category may occur as a single polygon or as a series of spatially disjunct polygons across the installation. Every occurrence of the various landcover/soil categories is considered a potential inventory site.

Depending on the amount of error inherent in the imagery and soil source data, and the error introduced operationally through data manipulation, geographic information system products may possess significant levels of error (3). Given this possibility of error, in addition to the minimum area required for the land condition field sampling method, it has been estimated that the landcover/soil polygons must be at least 2 hectares (5 acres) in size in order to be accurately identified and inventoried in the field. Therefore, the GRASS "macro" that recognizes the unique landcover/soil combinations has also been written in a form that eliminates all polygons that fail to meet this user-defined minimum area requirement.

An additional GRASS algorithm is used to randomly select polygons as field inventory sites. The number of selected polygons is dependent on the size of the military installation. The current policy is to allow one inventory site per 200 hectares (500 acres). For larger installations this may represent an unmanageable number of sites. Therefore, the maximum number of sample sites is limited to 200. These sites are randomly allocated to the polygons in a stratified fashion based on the percentage of the installation occupied by each landcover/soil category. This process ensures proportional representation of all landcover types and soil series. In addition, it allows the spectrally recognized landcover categories to be subdivided by soil series if field data indicate that more than one vegetation community occurs within a given landcover category.

Field crew leaders are provided with clear Mylar plastic overlays which correspond to US Geologic Survey quadrangle maps. The overlays are printed with all polygons of sufficient size to be sampled. The color scheme for the polygons is based on the landcover categories. Polygons selected for inventory by the randomization process are labeled with icons. Soil series delineations are outlined in black. It is the responsibility of the field crews to identify and inventory the areas represented by the

selected polygons. Once a given polygon is found, the field crew establishes a permanently marked vegetation transect that can be relocated and monitored in future years to evaluate trends of declining or improving land condition. In the event that any given polygon is inaccessible or unidentifiable in the field, the crew leader must select the next nearest polygon of the same landcover color code and soil series.

Conclusion

This procedure for land condition inventory site selection is currently being implemented at 15 major US Army training installations in the United States and West Germany. Future improvements in the procedure will depend largely on advancements in the field of remote sensing and image interpretation.

REFERENCES

(1) Lacey, R.M., W.D. Severinghaus and E.W. Novak. Integrated Training area management: the concept and strategy, p. 10. In: Agronomy Abstracts, 1988 Annual Meetings of the American Society of Agronomy, November 27-December 2, Anaheim, California.

(2) Diersing, V.E., R.B. Shaw, S.D. Warren and E.W. Novak. Methods for optimizing the management of U.S. Army training lands, p. 7. In: Agronomy Abstracts, 1986 Annual Meetings of the American Society of Agronomy, November 30 - December 5, New Orleans, Louisiana.

(3) Walsh, S.J., D.R. Lightfoot and D.R. Butler. Recognition and assessment of error in geographic information systems. Photogrammetric Engineering and Remote Sensing 53:1423-1430.

The Use of GRASS in Sampling Design - An Example

Susan C. F. Stitt

National Park Service
Geographic Information Systems Division
P.O. Box 25287
Denver, CO 80225

ABSTRACT

An example of designing a sampling scheme using GRASS GIS technology will be presented. Topics to be discussed include the use of GRASS tools in defining and redefining the population to be sampled and the need to clearly define the hypothesis being tested, as well as the benefits GIS technology can provide toward sampling design.

The Air Quality Division of the National Park Service was requested by the Pacific Northwest Regional Office to establish an baseline of elemental composition of vegetation and soils in North Cascades National Park Complex. This was to determine whether anthropogenic pollutants are being deposited in the park's ecosystems. The sampling process was to include analysis of subalpine fir, lichens, mosses, and soils. A sampling plan was to be developed using the existing geographic information system (GIS) data base for North Cascades National Park Complex. This data base includes a vegetation/cover type theme (Agee et al. 1985, Root et al. 1985), and elevation, slope, and aspect derived from Defense Mapping Agency (DMA) 1:250000 scale data.

Working with a member of the Air Quality Division staff and the team contracted to collect the field data, the sampling population was initially defined to be open canopy subalpine fir on south, southwest, and west facing slopes and within the park complex boundary. These aspects were chosen based on the assumption that the large air masses were flowing into the park from the southwest, and would appear in vegetation on southwest facing slopes earlier and more significantly than on other slopes. Open canopy was important because pollutants would presumably impact open canopy vegetation more easily than closed canopy vegetation.

The analysis to define areas fitting these restrictions was relatively simple using the GRASS GIS package. The vegetation data was masked to include only the area within the National Park Complex boundary. This layer was created as a cell file through Gmapcalc. This theme was then reclassed to include only open canopy subalpine fir and masked with south, southwest, and west facing slopes thus establishing the sampling population, or so we thought.

The next question was how to best sample this population. One of the criteria was that the sampling locations needed to be spread throughout the park complex to establish a parkwide baseline. It was thought that a total number of sites between 15 and 25 could be successfully sampled within a single field season.

Three possible strategies were proposed for choosing sampling locations. Simple random sampling, not chosen since it would not be likely to generate sampling locations spread throughout the park complex. The second proposed strategy was equal area / random sampling meaning splitting the park complex into regions contain-

ing equal park area from which random sites could be chosen. This method was not chosen since it would change the likelihood of any given cell being chosen as a sampling site. Areas with more sampling population cells would reduce the chance of any given one being chosen for sampling and areas with fewer sampling population cells would have an increased likelihood of any given one being chosen for sampling. The final proposed sampling method of equal population - random sampling seemed the most appropriate. It consisted of splitting the park complex into regions containing equal numbers of sampling population cells and then sampling randomly within each region. This method provided a means of spreading the sampling sites throughout the park complex without making any given cell any more or less likely to be chosen for sampling.

A report was run to determine the total number of sampling population cells, and it was decided to split the park complex into 25 regions within which a random sampling site would be selected. Initially the park was split into 5 regions running east/west each containing one fifth of the sampling population. This was accomplished by changing the window, then running a report to determine the number of sampling population cells within the window, and by trial and error, locating the window which would contain as close to one fifth the sampling population as possible. Again by trail and error, each east/west region was split into 5 smaller regions each containing one twenty-fifth of the total sampling population. Windowing in on each of the 25 rectangular regions, a single sampling point was randomly selected by generating random utm coordinate pairs through a computer driven random number generator, and using the first location which fell within a sampling population cell. This process was tedious, and required the generation of hundreds of coordinate pairs before locating one which fell within a sampling population cell.

The sites were plotted and the field crew visited the park, at which time the park staff reviewed the procedure and a few of the selected points were sampled. The accuracy of the sample locations being open canopy subalpine fir was low within the first few sample points visited, and the terrain was steep enough to make field work virtually impossible in some locations. In addition, the park staff requested that the sampling procedure be changed to allow for comparisons between watersheds. Their experience led them to believe the air was flowing in different patterns within different watershed regions of the park, the largest difference being between the areas east and west of the continental divide.

Seven watersheds (G. Larson et al., Oregon State Univ., Draft) were delineated on a park topographic map and then digitized and entered into the North Cascades data base. These then became the regions within which sampling was done and between which the sampling results could be compared.

The population to be sampled was redefined to ameliorate the problems encountered in the field. It was hypothesized that a given cell was more likely to be classified correctly as subalpine fir if it was not a single cell, but was within a large "polygon" of subalpine fir cells. A requirement was established that all stands of open canopy subalpine fir be at least 19 acres in size to be included in the sampling population. A cell file with these stands was generated by running Gclump on open canopy subalpine fir, then generating a report on this new layer, and manually choosing and reclassing only those clumps which were at least 19 acres in size. South, southwest, and west slopes were then used as a mask, and a new layer of open canopy subalpine fir on south, southwest or west facing slopes was developed. The problem of field work on extremely steep slopes was ameliorated by masking on slopes less than 65%.

It was requested by the field crew that at least 10 random sites within each watershed be generated, so that if a site was not correctly identified as subalpine fir, a new sampling location could be easily identified while in the field. At this time GRASS3.beta had been compiled on the computer being used, so the random sample locations within each watershed were generated through the new module random. These sites were plotted for the field sampling work.

The use of GIS technology greatly enhanced and streamlined the creation of a sampling design. The sampling strategy was changed in mid-project and a new sampling population was defined. This would have been more complex or perhaps even impossible without the use of a GIS. However much work could have been avoided if the question to be answered by this study, had been clearly defined earlier in the project. Specifically, are the comparisons of results to be done on an east-west and north-south basis, or between geographically defined regions such as watersheds. Through the use of GRASS GIS, the population to be sampled was readily defined and redefined within a short time frame, without a GIS, the same questions could not have been easily answered.

REFERENCES

Agee, J.K., S.G. Pickford, J. Kertis, M. Finney, R. de Gouvenain, S. Quinsey, M. Nyquist, R. Root, S. Stitt, G. Waggoner, and B. Titlow. 1985. Vegetation and fuel mapping of North Cascades National Park Service Complex, Final Report. NPS Contract CX-9000-3-E029. National Park Service Cooperative Studies Unit, College of Forest Resources, Univ. of Washington, Seattle, Washington. 111 pp. (map included).

Root, R.R., S.C.F. Stitt, M.O. Nyquist, G.S. Waggoner, and J.K. Agee. 1985. Vegetation and fire fuel models mapping of North Cascades National Park. Pages 287-294 in Pecora 10 Remote Sensing in Forest and Range Resource Management Proceedings. American Society for Photogrammetry and Remote Sensing.

Larson, G., G.A. Lomnicky, W.J. Liss. Classification of the Watersheds and Lakes of North Cascades National Park Service Complex. Oregon State University, Draft version.

THE APPLICATION OF GRASS-GIS IN ARCHEOLOGICAL INTRA-SITE SPATIAL ANALYSIS

by Ishmael Williams

Arkansas Archeological Survey
P.O. Box 1249
Fayetteville, AR 72702-1249

ABSTRACT

Archeological site data, like geographical data, consist of observations of the spatial properties of various phenomena, that can be manipulated and transformed to gain insights into problems of a multivariate nature. While, archeological site analysis operates on a much smaller geographic scale than most GIS-based studies, the multi-faceted complexity of a large excavated site containing thousands of artifacts, structural remains, and activity loci can pose as great a challenge in data management, analysis, and interpretation as any regional study. By applying a GIS system in concert with a relational database such as INFORMIX together with an exploratory data analysis system such as the S interactive statistical environment, this analytic task can be much improved over traditional archeological intra-site analysis procedures.

Background

Over the past two years the Arkansas Archeological Survey (AAS), with the cooperation of the U.S. Army Construction Engineering Research Lab (CERL), has engaged in developing GRASS-GIS applications in the assessment of site variability at Ft. Hood, Texas. Our most recent efforts to employ GRASS in archeological analysis focuses in on the individual site as the basic unit of study. This area of GRASS archeological applications at the AAS is only in its initial developmental stage and more work remains to be done before we can report fully on the utility of GRASS for this type of analysis. This paper discusses some of the characteristics of archeological site data to show how an approach that combines a relational database management system, a GIS, and a statistical system for exploratory data analysis can be implemented to improve the efficiency and flexibility of intra-site analysis.

Conventional Archeological Intra-Site Analysis

The archeologist seeks to reconstruct past lifeways by uncovering and bringing order to the distributions of tools, tool making debris, cooking hearths, trash pits, house support postholes, stockade lines, burials, food remains, and other deposits. Typically the archeologist has a number of aspects of site data to explore and multiple questions to answer. Some basic questions might relate to the vertical and horizontal locations and associations of well-datable diagnostic artifacts that can be used to determine the age and cultural affiliation of levels of the site or the existence of particular tool kits as inferred from the covariation over space of sets of artifacts. Beyond the standard goals of placing the site within a general temporal and cultural context, the archeologist might also want to explore the spatial aspects of site development in terms of the partition of the site into discrete use-areas corresponding to site occupation episodes, family household locations, task-

specific activity loci, group or public use-areas, family and village waste disposal areas, and ceremonial areas.

These levels of archeological pattern recognition must be inferred from basic units of artifact data recovery. This is accomplished through careful excavation of the deposits with very precise control over the provenience or vertical and horizontal location of archeological samples. Frequently there are two closely related components to a site occupation. The first is the consists of the narrow zone of archeological debris perhaps 10 to 30 cm thick extending horizontally across the site which encompasses the original living surface and contains most of the site debris. Such debris may include clusters of tools and pottery, lost or discarded artifacts, the waste debris from stone tool manufacture, and bits of shell, animal bone, and carbonized plant remains discarded as refuse. The second part of a site occupation includes the pits, house support postholes, stockade lines, burials, and hearths commonly referred to as features that were dug by the inhabitants of the site through the occupation level into the subsoil below. Data from the occupation surface are recovered through excavation of square sample units laid out across the site in a grid fashion while features usually are excavated as discrete units. Features and sampling units and the particular vertical levels in which each were excavated represent proveniences that constitute the fundamental units of site analysis.

After laboratory processing, conventional site analysis begins by examining the distribution of classes of artifacts such as pottery, lithic tools, lithic waste material, bone, plant remains and other samples for the site as a whole and for each of the feature and test unit proveniences. At some point in the analysis, the archeologist begins to focus in on the relationships between pairs or sets of multiple artifact classes that are associated with an activity or some other behavior that took place on the site as inferred from their covariation across features and test units and between different horizontal zones of the site. For lack of any other means, site data are often analyzed in a rote fashion to look for significant statistical trends in the univariate, bivariate, and multivariate associations of tool sets and artifact classes. In relying on cumbersome time-consuming batch programs, there is often little chance for multiple iterations to explore alternative ways of expressing the data and refining results.

To explore intra-site patterns, detailed site plans, consisting of meticulously hand-drawn maps of features, artifacts, and test units, are then manually or mentally overlaid to obtain a sense of the composition of the site with respect to the dozens of single, bivariate, and multivariate dimensions of the data. This sometimes includes the expression of raw or transformed data to assess spatial patterns in the distribution of artifacts using choroplethic mapping techniques. Because the techniques to accomplish these tasks are not well integrated, the process of setting up and running programs for intra-site spatial analysis consumes a large amount energy that could be better spent in the actual mental processes of site analysis.

A Comprehensive System for Intra-Site Analysis

What lacking in the conventional approach to site analysis is a comprehensive means of efficiently storing, displaying, combining, and manipulating artifact data in an interactive fashion to explore data and build a series of new site maps that derive from the resulting higher levels of understanding of the multidimensional aspects of the data attained at each step of the analysis. What is needed, in addition to access to a GIS like GRASS, is relational database management system to serve as a means of retrieving basic descriptive and locational information on artifacts referenced to each feature and sample unit provenience and linkage between the GIS and a compatible interactive exploratory data analysis (EDA) system. Fortunately, the components needed to build such a system are available now.

The first component, an efficient database, is met at the AAS with an INFORMIX relational database called DELOS developed by the Survey to access site level archeological data. DELOS is designed to afford flexible processing of data about archeological materials and their spatial context by linking

provenience information concerning the vertical and horizontal location of an artifact with descriptive observations about the morphology and cultural context of the artifact. The DELOS system for ordering artifacts is arranged in a hierarchical framework to allow for varying levels of specificity in the classification of artifacts. For example, one could access the locations, counts, and weights of all pottery from a site or pottery of a certain design and cultural affiliation, or only the counts of rim fragments for a certain pottery type or the vessel diameters for rims. In addition, these pottery attributes could be accessed for any or all discrete excavation units and vertical levels within units or for various pits, postmolds, and burials.

Two of the components of this site analysis system, a GIS and an EDA are already well integrated in GRASS (CERL 1988). GRASS, the Geographical Resources Analysis Support System, is a comprehensive Geographical Information System (GIS) developed for Army installations by the U.S. Army Construction Engineering Research Laboratory (CERL). GRASS is an integrated set of tools to manage land resources by providing means of inputting, storing, and manipulating data which are stored in maplayers consisting of spatially discrete cells across the region of interest. GRASS can store and process information in terms of a vectors (clumps of cells) or as coordinate point data (single cells). Many useful tools for intra-site analysis are available in GRASS such as mapping programs, nearest neighbor analysis, proximity analysis, cost-surface studies, coincidence and chi square tables, and many other boolean, mathematic, and algebraic functions that operate on mapsets.

The EDA component is met by "S", an interactive environment for data screening, analysis, and graphical display that runs under the UNIX operating system of Bell Laboratories (Becker and Chambers 1984). EDA is an inductive approach to searching for patterning in a dataset with the goal of of gaining insights into the nature of the data's total structure, particularly the unanticipated relationships that may occur. EDA involves iterative, stepwise examination and visual inspection of the many alternative representations of the data and utilizes graphic representations such as three-dimensional plot rotations of multivariate relationships to bring the brain's full visual processing capabilities into the gestalt of pattern recognition. S can also be used to setup deductive analysis by using the EDA capabilities as an entry level step in a multistage investigation where the relevant relationships are first assessed to explore the complex multidimensionality of the data prior to hypothesis formulation (Carr 1985).

In addition to the analytic tools provided within GRASS, GRASS is setup to transport information from datalayers to S via the GRASS to S module, and the S system is well setup for other sorts of analytic techniques that may be desirable in site analysis. Thus, if the mapset categories are features, sample units, and surface collection grids and the mapsets are all artifact classes recovered during the excavation, GRASS acts as a component in a database management system for setting up EDA in S. The AAS has developed modules in S that make available, in a menu form, macros for univariate, bivariate, and multivariate analysis of datasets that have been transported from GRASS. The modules will allow access to S statistical options such as boxplots, histograms, bivariate plots, regression analysis, three-dimensional spin of data swarms, cluster analysis, principal components analysis, discriminant function analysis, multidimensional analysis, and many others. There is also the capability of routing output from analysis in S back into GRASS as a new maplayer to be displayed and manipulated using GIS tools.

Setting up Site Data for GRASS and S Analysis

The AAS is in the process of analyzing and loading site data into GRASS for the Hardman Site (3Cl418) located in Clark County, Arkansas recently excavated by the AAS for the Arkansas Highway Transportation Department. Hardman is a prehistoric Caddo Indian habitation and salt processing site, dating between 1400 to about 1600 A.D. Excavations at Hardman recovered over 900 features identified as support-posts for houses, screens, and enclosures; refuse pits; burials; hearths; and thousands of

artifacts of stone, bone, and shell; fragmented and whole ceramic cooking, storage, and salt evaporation vessels; plant and animal food remains; and chronometric samples that will be used to date areas or levels at the site.

The excavation sample units of the occupation zone include a dozen 2 by 2 meter units and fifty-five 1 meter by 50 centimeter column samples. These units are being individually digitized as vector maplayers from records made in the field. Each unit vector has an identification label that corresponds to the field specimen number assigned all artifacts recovered within each level of each unit of the occupation zone. This maplayer of excavation units will be loaded with counts and weights of particular artifact types, bone, plant remains, etc., from the DELOS database and new seperate maplayers will be created for each class of archeological material. These datalayers will then be accessible to GRASS tools such as neighbors, Ginfer, and Gmapcalc to extrapolate patterns in the density distribution of artifacts. Datalayers can also be combined when appropriate with the overlay tools in GRASS to provide a view of patterns of multiple artifact distributions and evaluated with the coincidence tools to assess the association of pairs of artifact classes.

The features (hearths, burials, pits, and postholes) are being digitized as vectors or as coordinate points from site maps prepared from drawings and instrument readings made in the field. One maplayer will be made for each of the feature classes to permit flexibility in later segregation or overlay. The vectors are of a variety of shapes consisting of circles, ovals, and irregular amoeboid-like features which are being digitized using the stream mode. The majority of the postholes are being loaded as points which can be displayed with icons designed to approximate the circular shape and proportional diameter of the original posthole so that a realistic map of the site can be displayed. Like the excavation units artifact data from DELOS are being loaded into each feature vector to create many datalayers that can be analyzed with GRASS and S tools in a number of permutations of feature type and artifact class.

Since the vast number of postholes makes recognition of circular house patterns, enclosures, and other sets of related features difficult, the postholes maplayer will be subsetted to remove the noise created by multiple and overlapping occupation episodes. As we learn more about the site based on the archeological content of features, a series of maps will be generated to represent our understanding of the patterns and associations of pits, burials, postholes, and hearths and their affiliation with datable episodes of site use. Multivariate analysis of the artifact content of the occupation level and the features should help us to also obtain details about the spatial structure of the site with respect to the positions of activity use-areas which can also be displayed with the other site data to build a picture of the site for the various prehistoric occupation episodes.

The use of GRASS in combination with a relational database management system like DELOS and an interactive EDA such as S can go far in providing the tools necessary for fleshing out the multidimensional nature of site development and past lifeways. A detailed evaluation of the implementation of GRASS in the Hardman intra-site analysis will be reported later this year.

REFERENCES CITED

Becker, Richard A. and John M. Chambers 1984 An Interactive Environment for Data Analysis and Graphics. Wadsworth Advanced Book Program. Belmont, California.

Carr, Christopher 1985 For Concordance in Archeological Analysis, Bridging Data Structure, Quantitative Technique, and Theory. Westport Publishers in cooperation with the Institute for Quantitative Archaology, University of Arkansas. Kansas City.

CERL 1988 GRASS 3.0. September, 1988. United States Army Corps of Engineers, Construction Engineering Research Laboratory, Champaign, Illinois.

GRASS APPLICATIONS IN MACRO-SCALE CHOROPLETH MAPPING

J. J. Lockhart
Arkansas Archeological Survey
P.O. Box 1249
Fayetteville, AR 72701

ABSTRACT

The Arkansas Archeological Survey is currently preparing an overview of the cultural resources found in the U.S. Army Corps of Engineers' Southwestern Division. This paper outlines the processes involved in representing the spatially oriented attributes of the eight-state region using GRASS and the cartographic technique of choropleth mapping. In particular, the issues of data collection, standardization, classification, symbolization, and map production are discussed in terms of cartographic theory and GRASS applications.

Over the past several years, the Arkansas Archeological Survey has been preparing an overview of the cultural resources found in the U.S. Army Corps of Engineers' Southwestern Division. Among other tasks, the project involves the identification of cultural features found in the study area, and will result in a number of recommendations concerning resource management. These recommendations will be based, in part, on locational analysis made possible through the development of a database and corresponding GRASS data layers.

The study area encompasses almost 20 percent of the continental United States and is comprised of more than 600 counties in Arkansas, Louisiana, Texas, Oklahoma, New Mexico, and parts of Missouri, Kansas, and Colorado. The corresponding database for the Southwestern Division contains a number of individual data themes, ranging from various attributes of archeological interest to demographic information such as population density and change over time. Much of the available information was recorded in the form of county totals.

Data collected by statistical areas such as counties is often represented using the cartographic technique of choropleth mapping. From the Greek words, "choros" meaning place, and "plethos" meaning magnitude, the term "choropleth" denotes a specific type of representation in which quantitative thematic maps are used to symbolize the magnitude of ordinal level data within the boundaries of unit areas (Robinson et al., 1984).

The extensive use of choropleth maps may be due, in large part, to the efficiency with which they communicate geographic information, and the relative ease with which they can be produced (Anderson and Child, 1987). There are, however, several fundamental cartographic principles that should be considered in the design process if these maps are to be effective in terms of graphic communication. In particular, choropleth mapping is dependent on data collected by statistical or administrative areas such as states, counties, or census tracts. Because these units are often of unequal size, the data to be used is generally standardized such that it takes the form of some type of ratio such as densities or percentages. After the data is standardized, the data elements are typically grouped into four

to seven classes each of which is assigned a representative color or pattern. Using this technique, each area on a map is assigned the appropriate symbolization according to the data range it fits into. Finally, map elements such as the title, scale, legend, and data source are added to complete the cartographic product.

Data Collection The initial requirement for the mapping aspect of Southwestern Division Overview was the collection of quality information. The data used in the project comes from a variety of sources including the U.S. Census of Population, and archeological inventories for each of the six study units. Data for each attribute to be mapped was entered into a database according to county or county equivalent. So, the first data field consisted of the 609 county names sorted alphabetically. The next field contained each county's identification number. This number was also encoded into the corresponding county area on the base map in GRASS so that data values could be keyed to their locational counterparts. The next two fields consisted of state and study unit affiliation. Then, a field for each of the variables to be mapped was established, and the appropriate values for each county were entered.

Data Standardization Certain data fields, such as "county area in square miles" and "population by county", were added to the data base for purposes of standardization. As previously mentioned, data in the form of absolute, or raw, numbers alone are not ordinarily used in choropleth mapping (Robinson et al., 1984). Absolute numbers are typically standardized so that the data is represented either by ratios involving area such as densities, or ratios independent of area such as percentages or proportions. The reason for this is that most choropleth maps contain areas which are unequal in size. For example, Los Alamos County, a very small county in north-central New Mexico and its neighboring county, Santa Fe, have a similar number of archeological sites. However, to show them in the same class would be a misrepresentation due to fact that Santa Fe County is more than 17 times larger than Los Alamos County. To correct for this, archeological sites in each of the two counties were standardized by county area so that the finished map represented archeological sites per square mile. An example of standardization independent of area, on the other hand, might be a choroplethic representation of population change through time in which change for each county is shown as a percentage so that the unequal size of the counties is not a factor.

Data Classification After the Southwestern Division database was in place, the next step was to establish data categories for each attribute by grouping similar data elements into classes. The purpose for classing the data is to generalize, and thereby simplify and enhance the recognition of the geographic patterns. In order to maximize classing accuracy, areas which are quantitatively similar should be grouped together and represented by the same symbol. However, because of the existence of a variety of classing procedures, several different map distributions can be generated using a single data set. This situation poses a problem to cartographers concerning which classing method, or algorithm, to use with any given data set. While certain classing methods produce more accurate results with certain data distributions, some research has shown that the classing method most likely to produce accurate and reliable results with any distribution is the optimization procedure first proposed by Jenks and Caspall in 1971 (Smith, 1986). Optimization classing is an iterative process which establishes class intervals by minimizing variation within classes and maximizing variation between classes.

In addition to accurate classification, proper legend design can also enhance the effectiveness of choropleth map communication. The legend should contain the actual class limits without reporting values which do not occur in the data set. The result is to narrow the reader's estimate of the actual value of any area belonging to each category (Dent, 1985).

Data Symbolization Conceptually, by using area symbols with quantitative data, a statistical surface, or z-value, is implied. In choropleth mapping each statistical area is symbolized to represent the vertical height of its data value (Dent, 1985). This princi-

ple can be demonstrated in GRASS by assigning a vertical exaggeration to any choroplethic data layer, and then viewing the result from an oblique angle using the GRASS module "d.3d". A choropleth map, in other words, is a planimetric representation of a "stepped" statistical surface in that the data used is discrete rather than continuous, and distributions are controlled by political or administrative subdivisions.

The primary objective in cartographic symbolization is to preserve clarity and avoid visual confusion or ambiguity (Jenks and Knos, 1961). Ordinarily, since choropleth maps are intended to delineate areal differences in magnitude, class symbolization should be designed such that the reader can intuitively discern the hierarchic organization of the map categories. In other words, the reader should be able to determine the hierarchy even without a legend. Random or improper symbol selection, on the other hand, results in the reader continuously having to refer to the legend to determine the symbol hierarchy, which interferes with the communication of the geographic pattern being represented.

There are some cartographic conventions which relate to the gradation of colors for quantitative thematic mapping (Dent, 1985). The Simple Hue Plan is a one-color scheme which relies on variations in color value to achieve contrast between area symbols, and to establish the visual hierarchy. For example, symbolization would be comprised of a graded series ranging from a light, high-value hue for the representation of the lowest data elements to progressively darker shades of the same color as class limits increase.

Another gradation scheme is the Part-Spectral Plan which uses colors in the sequence in which they occur in the electromagnetic spectrum. Spectral colors range from violet, the shortest visible wavelength, consecutively through blues, greens, yellows, oranges, and reds, which are comprised of the longest visible wavelengths. The theory behind the Part-Spectral Plan is corroborated by a physiological phenomenon known as "advance-retreat" which holds that short wavelength colors focus in front of the retina and long wavelength colors focus behind. Consequently, the longer its wavelength, the closer a color will appear to the observer. The Part-Spectral Plan uses hue and value to differentiate between areal symbols. For example, classes could be symbolized by yellows, oranges, and reds, with yellow representing the lowest class limits and red representing the highest class limits.

Map Production Many of the maps to be used in illustrating the various data themes included in the COE Southwestern Division Overview were produced using GRASS software. An equal-area projection was chosen for the base map used in digitizing. Then, the map was registered using an arbitrary point of origin and meters as the map units. Each county was digitized, and encoded with the same number that it was represented by in the database file. After the encoding, the vectorized base map was converted into a cell file which became the base map for the reclassifications that produced the subsequent choropleth maps. After each data field was standardized, a sorted list of the values was run through a classing program to establish class limits for each map. Each county was then assigned to a class according to the individual data theme, and a script was generated and read into the GRASS "Greclass" module using the rasterized base map as the input data layer to create each new choropleth map. The maps were then supported with color and category information, and essential map elements such as title, legend, date, data source, and scale were added to complete the map production process.

In many analytical situations, being able to actually distinguish spatial patterns, rather than just looking at a column of data values, can be an important part of the decision-making or problem-solving process. The primary purpose in choropleth mapping is to portray the general distribution of an attribute. And, with adherence to a few basic cartographic design principles, choropleth maps can be valuable tools in the communication and interpretation of complex spatial relationships.

REFERENCES

Anderson, J., Child, J. 1987. Choropleth Maps: Window Dressing or Workhorses? GIS '87. Proceedings of the

Second Annual International Conference, Exhibits and Workshops on Geographic Information Systems. Falls Church, Virginia: American Society for Photogrammetry and Remote Sensing, and The American Congress on Surveying and Mapping. 352-361.

Dent, B.D. 1985. Principles of Thematic Mapping. Reading, Massachusetts. Addison-Wesley Publishing Company.

Jenks, G.F., Caspall, F.C. 1971. Error on Choropleth Maps: Definition, Measurement, Reduction. Annals, Association of American Geographers 61 (2): 217-244.

Jenks, G.F., Knos, D.S. 1961. The Use of Shaded Patterns in Graded Series. Annals, Association of American Geographers 51: 316-334.

Smith, R.M. 1986. Comparing Traditional Methods for Selecting Class Intervals on Choropleth Maps. Professional Geographer 38 (1): 62-67.

Robinson, A.H., Sale, R.D, Morrison, J.L., Muehrcke, P.C. 1984. Elements of Cartography. New York, New York. John Wiley and Sons, Inc.

RECLASSING IN GRASS AND ITS APPLICATIONS FOR SOIL ATTRIBUTE DATA

Pamela Thompson

U.S. Army Corp of Engineers
Construction Engineering Research Lab
P.O. Box 4005
Champaign, IL 61824-4005

ABSTRACT

The Geographical Resources Analysis and Support System (GRASS) is a a grid-cell based Geographical Information System (GIS). GRASS is a tool that can be used to manipulate map layers and perform analysis useful for environmental planners and land managers. Each map layer in GRASS is made up of two different types data: (1) the spatial data that designates where in space a particular geographic feature occurs, and (2) attribute data that assigns the geographic feature a specific label. New map layers can be created from existing layers by using the **RECLASS** function which assigns new attribute data to the existing spatial data.

Reclassing is especially useful for United States Department of Agriculture - Soil Conservation Service (USDA-SCS) soils maps, since each mapping unit on a soils map has several soil properties and interpretations associated with it. Map layers representing these properties and interpretations are useful for many types of analysis. In addition, USDA-SCS has their soil information entered into 2 databases which can be accessed through a search and retreival system, creating a readily available source for reclass information. However, SCS soil information is structured so that most of the data is based on soil taxonomic units and not on the mapping units depicted on a soils map. Reclassing the soil mapping units into properties and interpretations based on taxonomic units can become confusing, since mapping units often contain 2 or more taxonomic units. Furthermore, soil properties for any particular taxonomic unit are recorded by soil horizon. The depth increments for soil horizons vary from soil to soil, making reclassing for a specific property difficult. Because the reclassed maps derived from soils maps are useful and important for so many types of analysis, consideration must be given to the issues involved in reclassing soils maps.

RECLASSING IN GRASS

The Geographical Resources Analysis and Support System (GRASS) is a Geographical Information System (GIS) developed at the U.S. Army Construction Engineering Research Laboratory, Champaign, IL. GRASS is a tool for storing, combining, analyzing and displaying multiple map layers for use in environmental planning and land management. It is a grid-cell based GIS, but does have some vector display capabilities.

A map layer within GRASS is made up of two types of data; (1) the spatial data that designates where in space a particular geographic feature occurs, and (2) attribute data that assigns the geographic feature a specific label. For instance a vegetation map layer would consist of spatial data that puts the areas of specific vegetation types in the

correct position in space and attribute data that records what type of vegetation communities are present.

Since spatial and attribute data are entered and stored separately in GRASS, new map layers can be created from the existing spatial data simply by assigning a new set of attribute data. This is done by using the **reclass** or **Greclass** function in GRASS. In the above vegetation overlay example, the attribute data assigns each area of the map a vegetation type. This overlay can be reclassed to create a map representing any of a number of properties or characteristics of these vegetation types, such as cover density, total forage production, etc...

In the interest of saving disk space, reclass does not actually produce a new map, but instead, creates a reclass table containing the new attribute data which is stored and used to reclassify the original map layer whenever the new reclassed map name is requested. As far as the user is concerned, a new reclassed map has been created. Because reclass tables are based on the original map layer, reclass maps are only available as long as the original map layer remains in the database.

Reclassing is especially useful for creating additional maps from original soils maps. Soil maps are commonly reclassed into soil property or soil interpretation maps. These types of reclassed soil maps are useful as input for various analyses, such as evaluating soil erosion status, siting a new landfill or determining suitability for crop production. Another reason soils maps are particularly suitable for reclassing is that the United States Department of Agriculture - Soil Conservation Service has put their soils property and soil interpretation data into databases which can be accessed through a soils information system. Reclassing can be done using a soils information system and a data base management system (DBMS), or by entering reclass information directly.

SOILS DATA SOURCES

USDA-SCS (United States Department of Agriculture - Soil Conservation Service) is in the process of mapping soils for all areas of the United States. SCS produces soil survey reports, using nationally approved guidelines and definitions. These soil survey reports give general descriptions of mapping units, as well as estimates of soil properties such as texture, permeability, and have tables giving physical and chemical properties and use interpretations for the soil series and phases of a soil series. Soil survey reports are published for specific soil survey areas which are most commonly counties or groups of counties. These reports can be ordered by contacting either the SCS field office for the survey area of interest or by contacting the State SCS office for the state containing the survey area.

The soil information contained in soil survey reports is also available through SCS soil databases. The data developed in the process of making soil surveys are entered into a computer at the Statistical Laboratory, Iowa State University, Ames, Iowa.

The data are entered into two different databases, SOI-5 and SOI-6. The SOI-5 is the database for the taxonomic unit, usually soil series (and phases of soil series). It contains information from the Soil Interpretation Record, which consists of a brief soil description, as well as estimates of soil properties such as texture, permeability, depth to bedrock, frequency and duration of flooding, yield estimates of crops, woodland and range production under stated management systems, suitability or limitations of soils for specified land uses, and soil features affecting specified land uses.

The SOI-6 database is the database for the mapping unit. It contains information from the Map Unit Records, which consists of mapping unit characteristics (such as slope, USDA texture, flooding frequency, prime farmland code), critical phase criteria, and survey acreage by county. The SOI-6 database does not, however, contain information about specific properties (other than those listed above) or about use interpretations. This type of information is only in the SOI-5 database.

It is important to distinguish between taxa in soil classification and mapping units on a soil map. *Soil Taxonomy* (Soil Survey Staff, 1975), makes the distinction between the taxonomic unit (or soil series, family, great group, etc.) and the units shown on

the map (map units), saying they are two different things and must not be confused, even if they carry the same name. Taxonomic units are described and defined using clear rules and guidelines. However, because soils can vary greatly in their properties, they rarely fit neatly into the bounds of the taxonomy. The mapping units often represent variations from the central concept of the taxonomic unit. In addition, mapping units are used when separating individual soil taxonomic units would be impractical. For large-scale maps (SCS defines large-scale maps as maps with scales of 1:31,680 or larger), mapping units are most commonly named as phases of a soil series. But, soils can often be too small or too intricately associated with other soils to mapped separately at scales commonly used for soil maps. In these cases, mapping units are commonly named as a complex of two or more soil series, an association of two or more series, or a combination of two or more undifferentiated groups, depending on the variability of the delineated map areas and on the constraints placed by the scale of the map.

APPROACHES TO RECLASSING SOILS MAPS

Information for reclassing soils maps can be acquired from soil survey reports or from soil databases, such as the SOI-5 and SOI-6 databases. Whether using a soil survey report directly or using soil databases, there are inherent problems trying to relate the conceptual taxonomic unit (SOI-5) to the soil mapping unit (SOI-6) for the purposes of reclassing soil maps for further analysis. It becomes difficult to assign specific soil properties or interpretations to a soil mapping unit that sometimes includes more than one taxonomic unit. In particular, soil complexes present a problem because they can consist of component soil series that can vary greatly in their properties or use interpretations. Soil associations do not seem to cause a problem, since they are usually mapped only when their component soils are similar in their properties and use interpretations. In a soil survey report, soil complexes are described with the approximate relative percentages of its component soils included. When using a soil survey report to reclass, soil complexes could be named with the percentages of component soils included. Then, when using the reclassed map for further analysis, a weighted average approach might be taken. Currently, the SOI-6 database (Map Unit Use File) does not give relative percentages of component soils for soil complexes.

One way of approaching this problem of reclassing soil complexes using the SCS databases is to reclass using the values from the SOI-5 database for the element of the complex that constitutes the greatest area of the mapping unit. Complexes are named by listing the name of the soil that makes up the greatest portion of the mapping area first. If the goal for reclassing is to create a general soil properties map, this approach may be sufficient. However, when the map is to be used for specific management purposes, such as suitability for pesticide application or suitability for landfill cover, this method may result in classifying an area as suitable when somewhere within the area there are soils with severe limitations.

Another way of reclassing soil maps is to use a limiting factor approach. For this method, soil complexes or associations (if needed) would be assigned values corresponding to the soil element that has the most limiting factor for the purposes of the map. Of course, this assumes that the uses for the map are known before reclassing. For instance, if a soil map is being reclassed to a map showing suitability for basements with dwellings, the soil complexes would be given values corresponding to the soil element within them that has the most severe limitation for dwelling with basements. This would prevent the area mapped as a complex from being classified as good for dwellings with basements when it may actually contain a soil that has very high shrink-swell. When using the limiting factor approach for reclassing a soil map to soil properties such as pH, texture, bulk density or erosion factor, the ultimate use for the map must be known in order to reclass using the correct value (high or low values being limiting?). For instance, if a soils map is being reclassed to represent soil permeability and the resulting map is to be used to indicate those soils suitable for

landfill construction, soil complexes should be assigned values corresponding to the soil element in the complex that has the highest permeability value. This is done because soils with high permeabilities have limitations for landfill construction, since waste materials may leach through the soil and reach the ground water quickly and easily. If the same soils map is being reclassed to represent soil permeability for the purpose of indicating soils with limitations to septic tank construction, just the opposite approach for reclassing should be used. Soils with low permeability values are not useful for septic tank construction.

The concept of soil horizons also creates a concern when reclassing soils maps into soil properties or use interpretations. Soils are defined, in part, by the character and thickness of their soil horizons. In the SOI-5 database, soil properties are reported in tables for each soil horizon. The depth of these soil horizons) varies from soil series to soil series. For example, the soil surface may be the first 5 inches for one soil and the first 24 inches for the other. The second horizon may be 6-10 inches for one and 9-47 inches for the other. To get around this problem, general surface and subsurface categories can be used. For example layer 1= first horizon, regardless if its 0-5 inches or 0-24 inches. Four layers should be sufficient for most soils; (1) surface horizon, (2) second horizon, (3) third horizon, and (4) below the third horizon.

SOILS RECLASS EXAMPLES

The following are some of the more common soil properties, use interpretations and land use suitability classifications that may be of interest to land managers. This type of soil information could be used to reclass an original soils map. A suggested number of layers to be reclassed are given for each soil property. When reclassing soil complexes and associations, an example on how to use the limiting factor approach is also given for each soil property and use interpretation. The value ranges, where appropriate, correspond with those set up by SCS and used in the SOI-5 database. These value ranges may be used for categories in the reclass map.

1. USDA texture Limiting Factor-use the soils with silty textures. Map may be used for soil erosion analysis, silty soils erode more easily.

2. percent organic matter Limiting Factor-use the soil with the lowest % organic matter. Map may be used to indicate which soils have poor structure and therefore easily compacted or eroded. Scheduling maneuver training on these soils should avoided in wet conditions.

   > = 0 but < 1
   > = 1 but < 2
   > = 2 but < 5
   > = 5 but < 20
   > = 20

3. permeability (minimum in/hr) Limiting Factor-create two reclassed maps, one using the soil with the highest permeability and one using soils with the lowest. Low permeability may be favorable for analysis concerning landfill cover, while high permeability may be useful for septic tank construction. Low permeability may also indicate wet conditions during periods of high precipitaiton. Maneuver training may be complicated by wet, muddy soils.

   > = 0 but < .06
   > = .06 but < .2
   > = .2 but < .6
   > = .6 but < 2.0
   > = 2.0 but < 6.0
   > = 6.0 but < 20.0

4. available water capacity (total inches) Limiting Factor-use soils with low available water. Map may be used to indicate soils with limitations for revegetation. Low available water would be detrimental for plant growth.

   > = 0 but < 3
   > = 3 but < 4
   > = 4 but < 5
   > = 5 but < 6
   > = 6

5. erosion factors (K and T) Limiting Factor-use soils with the highest K factor and the highest T factor. Map may be used for erosion analysis. Low K factors mean the soil is easily eroded. High T values mean

6. moist bulk density (maximum g/cc) Limiting Factor-use the soils with the highest bulk densities. Map may be used to indicate soils with limitations for revegetation. High bulk densities can be detrimental for plant root growth. High bulk densities, however may be desirable for road building.

   > = 0 but < 1.0
   > = 1.0 but < 1.2
   > = 1.2 but < 1.4
   > = 1.4 but < 1.6
   > = 1.6 but < 1.8
   > = 1.8

7. pH Limiting Factor-use soils with the lowest pH. Map may be used to indicate soils with limitations for revegetation. Low pH is generally detrimental for plant growth.

   > = 0 but < 3.6
   > = 3.6 but < 4.5
   > = 4.5 but < 5.6
   > = 5.6 but < 6.6
   > = 6.6 but < 7.4
   > = 7.4 but < 8.5
   > = 8.5

8. salinity Limiting Factor-use soils with the highest salinity. Map may be used for to indicate soils with limitations for revegetation. High salt content can generally be detrimental for plant growth.

   > = 0 but < 2
   > = 2 but < 4
   > = 4 but < 8
   > = 8 but < 16
   > = 16

9. flooding and high water table - includes the following information: Limiting Factor-use the soils with the most frequent flooding and the most shallow high water table depth. These soils could present problems during periods of high precipitation.

   a. flooding frequency.
   b. flooding duration.
   c. flooding months.
   d. high water table depth
   e. water table kind.
   f. high water table months.

soil interpretations and land use suitability

Limiting Factor-use the soils with the most severe limitations for the following interpretations.

1. sanitary facilities - includes ratings for the following:

a. septic tank absorption fields.
b. sewage lagoons.
c. sanitary landfill (trench)
d. sanitary landfill (area)
e. daily cover for landfill.

2. water management - includes ratings for the following:

a. pond reservoir area
b. embankments, dikes and levees
c. excavated ponds - aquifer fed
d. drainage
e. irrigation
f. terraces and diversions
g. grassed waterways

3. wildlife habitat suitability - includes the following information:

a. Potential for several habitat elements.
b. Overall potential for:

1. openland wildlife.
2. woodland wildlife.
3. wetland wildlife.
4. rangeland wildlife.

Glossary

The following are definitions that will be useful for this discussion about soil data and reclassing in GRASS.

(1) Taxonomic Unit - A named kind of soils(taxon) that has specific properties with defined limits or ranges in characteristics. Each class within the six categories of "Soil

Taxonomy" is a taxonomic unit.

(2) Soil Series - A group of soils having horizons that are similar in differentiating characteristics, except for differences in texture of the surface layer or of the underlying material. All soils of a soils series have major horizons that are similar in composition, thickness and arrangement.

(3) Phase of a Taxonomic Unit - A subdivision of a taxon based on texture, stoniness, erosion, salinity, contrasting substratum, etc. Generally used in combination with descriptive terms that define the slope, physiographic position, or special environmental characteristics of the map unit. A phase bridges the gap between the taxon and the map unit.

(4) Phase of a Soil Series - A subdivision based on one or more characteristics that are potentially significant to use or management of the soil. The most common basis for delineating phases is slope, surface texture, erosion, stoniness, salinity, contrasting substratum, physiographic position, and flooding frequency.

(5) Soil Map Unit - An area of soil(s) delineated on a soil map. It contains one or more taxonomic units.

(6) Consociation - A map unit that is dominated by a single kind of soils of miscellaneous area.

(7) Undifferentiated Group - Two or more taxonomic units that are not regularly associated together. The members of an undifferentiated group commonly are similar enough in morphology and/or behavior so that separating them on the map is not important for the objective of the survey. Such a unit is named by combining the names of the taxonomic units with "and".

(8) Soil Complex - Two or more taxonomic units that occur together in a more or less regular pattern and are so intricately mixed, or so small in size, that it is not practical to separate them in mapping. The members of a complex commonly have contrasting morphology, as well as potentially unique use or management, but cannot be separated at the map scale being used.

(9) Soil Association - An association is similar to a soil complex except the members of an association could be separated at scales commonly used on detailed soil maps (15,840 - 24,000).

10) Soil Horizon - A soil horizon is a layer of soil approximately parallel to the soil surface with characteristics influenced by genetic processes. Each horizon is separated from adjacent ones on the basis of differences in properties. The composition and arrangement of soil horizons in a soil profile (a vertical cut exposing the various parts of a soil) are the major determinants in the classification, mapping and use of land areas.

REFERENCES

Soil Taxonomy, a basic sytem of soil classification for making and interpretating soil surveys, USDA Agriculture Handbook 436, 1975, Washington, DC., V.S. Government Printing Office, pp. 754.

An Interactive Soils Information System User's Manual, U.S. Army Corp of Engineers, Construction Engineering Research Lab, Pamela Thompson, Keith Young, William D. Goran, Al Moy. Technical Report N-87/18, July 1987, 83 pages.

FIXING ARTIFACTS IN THE DATA YOU MUST WORK WITH (PART 1): USING GRASS TO INSPECT AND EDIT DIGITAL ELEVATION MODELS (Initiating a new series (?) on data quality control using GIS)

David Hastings

NOAA
National Geophysical Data Center
325 Broadway
Boulder, CO 80303

ABSTRACT

Digital elevation models are fundamental to many Geographic Information Systems (GIS) projects. Data sets such as slope, aspect, shaded relief, and watershed models derived from DEMs are also important to many projects.

However, none of the digital elevation models currently available have been produced with your application in mind. Most of these data are characterized by artifacts that can adversely affect your project. Many of these artifacts can be detected and at least partially alleviated in a GIS or image processing system. Another advantage of such processing is the possibility of completely documenting the work.

The Pinion Canyon are, near Trinidad and La Junta in southeastern Colorado, is used to illustrate the detection and partial repair of DEM data in GRASS. Although the National Park Service systemused for this work was performing beta-tests of GRASS 3.0, this study used GRASS 2.0.

Simple color display may not adequately detect artifacts (though in Pinion Canyon many artifacts are so obvious that the simplest of visual inspections in a GIS will detect them); more sophisticated but simple-to-produce displays using very tight color density slices, computations of slope and aspect often show such features of data.

Simple data dropouts, for which the values of individual grid cells at the edges of quadrangles are zero, can be repaired by combinations of filtering and patching. More complex dropouts, where the values may be almost (but not quite) zero, or where they may be unrealistically high (3700 meters at sutures of the mosaiced quadrangles in some parts of Pinion Canyon, where true elevations are 1200-1800 meters) cannot be corrected so simply, as patch will only arbitrarily replace zero values, rather than user-assigned values of a map. In Pinion Canyon, a binary mask of areas with unrealistic values was created using "rescale," which was then used to reassign bad values to zero for repair by patching.

"Gmfilter" was used to create custom filtes that reduced the patchiness of the DEM data. Aspect is a very unforgiving display of artifacts in DEMs and was used to evaluate the results of several "Gmfilter" windows.

Various versions of the artifacts in the Pinion Canyon DEMs are shown in detail, with discussions of a number of options for their repair. More sophisticated filtering and neighborhood analyses would help GRASS to better perform quality control functions. Despite being incomplete, current GRASS capabilities help the user to document/improve a database.

Digital elevation models are fundamental to many Geographic Information System (GIS) projects. Data sets such as slope, aspect, shaded relief, and watershed models derived from DEMs are also important to many projects.

However, none of the digital elevation models available from the U. S. Geological Survey (USGS), the National Oceanic and Atmospheric Administration (NOAA), or the Defense Mapping Agency (DMA) were designed with your GIS application in mind. Indeed, the skeptic would say that each data set is characterized by artifacts, with actual information being somewhat secondary. For example: 1.) Early digital elevation models (DEMs) were created to help produce molds for the plastic relief maps sold by the DMA in the 1960s. It is not surprising that, when someone tried to use them analytically, s/he found some undesirable characteristics, such as concentrations of digital values around the contours on the paper maps. Compute slopes from such data and you get near-zero values almost everywhere but for areas mid-way between contours on the paper maps, where the "roundoff" from one contour value to the next occurs in the digital data. Not the best way to model floodplains around your prospective dam site.

2.) Some DEMs are produced by digitizing with one line spacing, then resampling to a closer or a coarser spacing. With one type of DEM, values are digitized along lines separated by 90 meters, then resampled to 30-meter grid spacings. One can see the cost justification, but what physical justification allows this? These are not geophysical potential field data, where the value at any point is a simple function of the values at all other points in space. One could virtually digitize the terrain around Devil's Tower (National Monument, Wyoming), yet miss the tower itself in data represented to have a 30-meter grid spacing! Such data could be a real nightmare for GIS processing of a highway design. (Incidentally, there does not appear to be a way of repairing such data. They are best thrown out [or perhaps resampled to more appropriate 100-m spacings]).

3. Often considered the "premium" DEMs in the 7 1/2 minute USGS series, data from the Gestalt Photomapper are produced in 500 meter square "patches" from aerial photographic stereo pairs. These patches are then mosaiced with inadequate horizontal- vertical control to produce the models. The resultant data are often inappropriate for contouring, let alone deriving such data as slope and aspect.

4. Global data sets produced by NOAA's National Geophysical Data Center are combinations of regional or discipline-specific data sets produced by other laboratories, often for disparate interests. Land values may be digitized from maps (or estimated where the maps have no values) by a meteorological laboratory. Bathymetric models created by another lab by digitally interpolating actual bathymetric soundings combined with modelled values based on concepts of the shape of the ocean floor. Grid sizes may be different; values on one grid may be based on grid centers, others on grid nodes. Combinations of these data sets may be valuable interim data sets for global modeling. But care should be taken in making these models.

5. Attempts to make DEM data appropriate for mosaicing quadrangles into larger study areas have not been completely successful. Bathymetric models may use different representations of coastlines from models of land elevations. 3 arc-second DEMs sometimes have significant vertical discontinuities at section-lines, while mosaics of 7 1/2 minute DEMs (which have no overlap [sometimes

at the expense of having data gaps] at quadrangle edges) can have line or column dropouts, or even slivers of extremely high values at sutures between quadrangles.

Before I sound overly critical, let me say that current efforts to produce digital data sets are truly pioneering. These pioneers are forging up the digital Missouri river in their digital canoes, never imagining what kinds of Kansas Cities we will be creating from their efforts. Nevertheless, while we continue to try to create our Kansas Cities from such pioneering efforts, we should try to avoid building roads that slide into canyons, bridges or dams that collapse or flood unexpected areas, or dump radioactive waste (unexpectedly) upstream from a major aquifer. Inspection and partial repair of data provided to us by others may help us to avoid such callamities.

IMPROVING THE QUALITY OF DIGITAL ELEVATION MODELS USING GIS

The convenient display capabilities of a GIS or image processing system allows the analyst to inspect data in a variety of ways. Grey-shade images, color displays using continuous rainbow colors, color density slices using "random" colors for many small ranges of the data, shaded-relief images (GRASS aspect images are often easy-to-compute approximations of shaded-relief), slope images, combinations with other data, etc., all have unique capabilities to enhance certain artifacts in a data set.

For example, your data may have been inappropriately supplied to you as eight-bit (0-255) or integer values, when they should have been supplied as decimal fractions or real values. Try to compute slope from gravity data provided as integers, and see the "flat" areas interspersed by "cliffs" where digital roundoff occurs in the input data. This is an inappropriate representation of the data. You may be able to make limited use of such data, but they may be inappropriate for your main application.

In our specific case, production of slope maps quickly helps the analyst to spot major data discontinuities at sutures between quadrangles. Very high slopes along quadrangle boundaries appear for either data drop-outs or slivers of erroneously high values. "Random" color density slice displays of elevation data should look a bit like psychedelically colored topographic contour maps. Each color slice should follow the terrain in a physiographically realistic pattern. If there are rectangular grid-like deviations, if there are linear discontinuities in the patterns, you probably have bad data. Produce an aspect image. Artifacts should be accentuated in such a display.

The repair of the DEMs can be divided into in two forms:

1. repair of suture lines of mosaics, and

2. as much as possible, removing the artifacts resulting from the specific data production methods in each quadrangle, that reduce the usefulness of the data for further processing for ones application.

Few Geographic Information Systems have the complete functionality to handle errors in the data. GRASS is not yet mature enough to completely handle the errors that can be repaired by the user. (Remember, it is almost always better to have had the data brought to as high a standard by the producers of such data - and your encouragement of such efforts by data producers may be in mankind's interest as well as your own.)

But GRASS, as well as many other GISs and image processing systems, has patching capabilities to partially repair bad data values, mosaicing capabilities to remosaic quadrangles, filtering capabilities to subdue the effects of Gestalt Photomapper patches, and so forth. These functions can serve to help us improve the quality of DEMs and other data. Ultimately, expert systems will be developed that will directly query specific types of data (rather than a human interface between the expert system and the data that both reside on the same computer as is currently often the case). Pattern recognition (machine vision) techniques will detect and automatically document and alleviate the most common types of data

errors.

USING ONE EXAMPLE TO ILLUSTRATE THE STYLE OF REPAIR TECHNIQUES FOR ONE TYPE OF DIGITAL ELEVATION MODEL

The Pinion Canyon area between Trinidad and La Junta in southeastern Colorado is used as an example of the procedures that can be used in GRASS to improve the usefulness of one type of DEM for further GIS processing. Pinion Canyon is recently acquired Army land, part of Fort Carson. Prior to acquisition, it was ranchland. A Grass GIS data base is being constructed to help manage the environmental resources of the land. The author is investigating the use of GRASS to produce hydrologic models in the area. The work is being done on the MassComp 5600 - based GRASS system at the Geographic Information Systems Division of the National Park Service's offices in Lakewood, Colorado.

Digital Elevation Models are a fundamental part of this GRASS data base. As is typical of a GIS exercise, the data were not originally produced with GIS applications (let alone the specific applications needed by the Army Corps of Engineers Construction Engineering Research Laboratory or Fort Carson management). The data should be inspected for characteristics that might affect GIS processing. It is also worth alleviating whatever negative artifacts the data may have for a particular application. This step in data base development is often omitted, to the detriment of the users' objectives.

The digital elevation models of Pinion Canyon obtained from the U. S. Geological Survey were produced on the Gestalt Photomapper. This device works directly with aerial photographic stereo pairs to produce models within individual "patches" (almost square rectangular areas). Several of these patches are then mosaiced without accurate vertical control to produce the DEM for a particular 7 1/2 minute quadrangle. The 7-meter accuracy described for such DEMs is calculated by comparing values at section lines where adjacent quadrangles have DEMs. There is no comparison with actual locations of benchmarks during this accuracy assessment, and the sharp variations in elevation at the edges of the patches are essentially overlooked by the producers of the data.

In addition, it is a policy of the producers of these data to avoid overlaps in data coverage at the edges of quadrangles. Due to the nature of the Universal Transverse Mercator projection's fitting of a flat surface to the Earth's curviture and the nature of production of the DEM data, this policy results in slivers of data drop-out along sutures when these quadrangles are mosaiced. These data dropouts can have zero value, very high values (much higher than any physical elevation in the area), or something in- between.

The patches on the Pinion Canyon data are very disturbing, not unusual for such DEMs. Such artifacts are clear evidence that the data are not produced for rigorous analysis, such as pattern recognition or the computation of slope or topographic aspect. In some cases, even casual inspection of the raw elevations is disturbing, let alone rigorous modeling of derived data sets (such as slope) in a GIS. The latter may produce an outright fallacious result without extreme caution in the GIS processing.

Many scientists are reluctant to "tamper" with the DEMs, preferring to accept the artifacts as given. But the data were originally produced with a somewhat arbitrary procedure. When one realizes that the data are digitized on one unevenly spaced grid, then resampled to another grid (without any physical justification - we are not dealing with potential fields here!), and that neighboring (rather than nonexisting overlapping) values are statistically compared to check on the vertical precision of the data, we see that the producers of the data are using physically misleading (though statistically "valid?") methods to claim their 7-meter accuracy. We should understand our own objectives in using these data, and the conflict between the original production methods and our objectives. With this in mind, we should feel no reluctance about reworking the data to make them more appropriate for our specific needs.

In the Pinion Canyon area, the sharp changes in elevation at the edges of the Gestalt Photomapper patches result in inaccurately high values of slope, and inaccurate changes in value of aspect. The patching is poorly controlled.

In a hydrologic model we are less concerned with the overall vertical precision than with the relative distribution of elevation. Applying spatial filtering techniques may alter the absolute values of elevation, while locally improving the relative values.

Initial evaluations of the use of spatial filters to improve the DEMs consisted of repeated applications of the GRASS function "neighbors" to the display version of the DEM (ELEV.DEM). 3x3 mean filtering was applied repeatedly. One could apply the random color lookup table to the raw elevation data and have some trouble recognizing intuitively logical landforms in the data. Repeated application of the mean 3x3 mean filter led to increasingly physiographically realistic renditions of the area.

With the positive result of this initial test, the function "Gmfilter" was used to filter the actual DEM values (ELEV.DEM.TRUE). "Gmfilter" has the disadvantage of loading the entire map into memory. Not only this, but if one runs the function in "parallel" mode (to avoid corrupting the input data by previous processing) one needs to have both input and output data in memory. With the 4 megabytes in the National Park Service's MASSCOMP 5600, only about 1/4 of the Pinion Canyon mapset can be processed at once, thus requiring subscening before processing, with subsequent mosaicking to recreate the entire mapset.

Experimentation with various filter sizes included 3x3, 5x5, and 7x7 filters with different combinations of weights.

First, 1x3 and 1x5 vertical and horizontal filters were applied to the data, to see if symmetrical X-Y filtering was appropriate, or if different sizes and/or weights were needed in the horizontal and vertical directions. Gmfilter was run with filter weights such as the following:

```
0 1 0   0 0 0   0 0 1 0 0   0 0 0 0 0
0 1 0   1 1 1   0 0 1 0 0   0 0 0 0 0
0 1 0   0 0 0   0 0 1 0 0   1 1 1 1 1
                0 0 1 0 0   0 0 0 0 0
                0 0 1 0 0   0 0 0 0 0
```

The 3x3 filter needed to be run far more times than the 5x5 filter to produce acceptable smoothing. It was also found that the data were sufficiently symmetrical in the horizontal and vertical directions to permit the horizontal filters to be combined:

```
0 1 0   0 0 1 0 0
1 1 1   0 0 1 0 0
0 1 0   1 1 1 1 1
        0 0 1 0 0
        0 0 1 0 0
```

Applying such filters greatly improved the visual the elevation data. However, there was still an unrealistically grid-like pattern in the slope and aspect data derived (using Gslope_aspect) from such filtered elevation data. Thus hybrid filters were produced that allowed diagonally positioned elevations to influence the result:

```
0 0 1 0 0
0 1 2 1 0
1 2 2 2 1
0 1 2 1 0
0 0 1 0 0
```

Notice that the filters now include weighting favoring closer values over more distant values. This should be physically valid, though it probably is not particularly more valid than equal weighting of the data as actually produced by USGS.

Such filtering reduced the grid-like appearance of the slope and aspect data calculated from such filtered DEMs. It was decided, however, to experiment with increased filter size to 7x7, as well as with increased weighting along diagonals:

```
1 1 1 1 1 1 1
1 1 1 2 1 1 1
1 1 2 3 2 1 1
1 2 3 3 3 2 1
1 1 2 3 2 1 1
1 1 1 2 1 1 1
1 1 1 1 1 1 1
```

This filter kernel is now being used to process the data (ELEV.DEM.TRUE) for northwestern, northeastern,

southeastern, and southwestern quadrants of the Pinion Canyon mapset. Current assessment of the data is that funning this filter in parallel through the data fewer than four times leaves too much patchwork gridding in the data, and that more than four applications of the filter smoothes out too much detail. Four applications leave a combination of these problems: somewhat too much patchwork gridding in some areas, somewhat more smoothing than desired in other areas. But this compromise appears to be the best for the Pinion Canyon data.

THIS PAPER IS DISTRIBUTED ON A USER-BEWARE BASIS. IT FAILS TO COMPLETELY DISCUSS FILTERING AND PATCHING OF BAD MOSAIC SUTURES.

DATA CONCERNS

Versatility of GRASS Data Layers

Robert C. Lozar

Construction Engineering Research Laboratory
P.O. Box 4005, Champaign, Il 61824

ABSTRACT

The Geographical Resource Analysis Support System (GRASS) has been implemented at several installations across the U.S. Each implementation requires the development of a set of supporting data layers. Over the last several years, CERL has supported its implementation by the development of these layers. This is a retrospective of the work done for several installations. The paper reviews which layers were generated, how they were used, how often they were used and for what purposes. Based on which data set configurations have in the past proven to be the most versatile the paper identifies which layers are likely to provide the greatest return on the development moneys invested and how to make these decisions.

One of the major components of the establishment in a Geographical Resource Analysis Support System (GRASS) Geographical Information System (GIS), which data to generate to support the usage of the system is an important decision. Maps (called data layers) which are stored in the system are often expensive to translate from paper (or digital) form into the format used by GRASS. In an environment where budgets are limited, it is necessary to set priorities on which data layers can be generated. The layers which must receive the highest priority will depend on the applications to which the GRASS system is intended to be put. Since applications vary depending on location and agency, so must the desired map layers.

CERL has now had considerable experience in developing a set of initial data layers at several locations. Also CERL has been involved in carrying out a variety of applications using these layers. This paper will review some of the experiences which have been gained and make a set of recommendations based on those experiences.

Beside simply deciding on a group of layers to support a single application, several corollary questions need also to be asked. If you have identified a set of layers to be developed for a particular purpose, are there other applications to which this set can be put so the data layer development process will provide enhanced value to the final product? And if a layer is developed, can the categories easily be translated to another layer (e.g. using the GRASS tool called *RECLASS*) such that you have developed two or more layers for the cost of one? A notable example of this would be generating a map of soil "Ph" from the published Soil Conservation Survey County Soils Survey report. Choosing the initial data layer configuration carefully with these considerations in mind can considerably enhance the value and versatility of your GRASS data base.

There are several ways of determining a configuration of complementary data layers. Often this has been based on previous experience with data versatility and professional judgement. Topographic elevation might be translated into GRASS format, not because it is important (elevation often is however) but because elevation information can, with relative ease, be translated into

other layers which have great versatility. Elevation data is regularly translated into a layer showing the degree of slope. This may be used for modeling erosion potentials. Elevation may also be translated into a layer showing the aspect (direction in which the sloping land faces). This may be used in modeling archeological site occurrence potential. These types of considerations are important in setting priorities for data layer development.

Table I is an *initial response to the* question, "If one were to develop a general list of applications versus data layers, what would such a list look like?" Table I shows such an arrangement of potential applications arranged down the left column and data layers arranged across the top. Is it reasonable that a data layer will be useful in carrying out one of the applications? (The applications listed relate to CERL's experiences in developing layers for military purposes). At each intersection two responses are noted: 1. This data layer is normally a required input to carry out the application under consideration or 2. This data layer would be helpful in carrying out the application under consideration but the modeling can be carried out without its presence. From a matrix like this, we can derive an understanding of which data layers potentially have the highest versatility in supporting different applications.

From Table I, it is clear that a few layers stand out for their versatility in different applications. Satellite or digitally stored remote sensed images (e.g. National High Altitude Photography- NHAP) have a high degree of versatility largely because the color or spectral data they contain can be reinterpreted to indicate land cover or land use type. (i.e. They may be used to generate a type of vegetation map, or they may be used to imply cultural features such as urbanized areas.) Remotely sensed images also provide a historical document of the changes which have occurred in an area. Another example are soils data. Soils data are versatile because their supporting reports usually correlate the distribution of the soil with a wealth of information about the characteristics that the soil type imply (e.g. its engineering properties, its fertility, the natural history of its formation). Another, topographic elevation, is useful (as mentioned before) because from it, it is easy to generate slope and aspect data as well.

The usefulness of this chart comes from realizing that if your data layers include slope, soils, landcover and roads developed initially to support the application of erosion control, you have the set of information required to begin to deal with questions about forestry applications.

At the bottom of this matrix are two rows which characterize 1. the difficulty of generating the data layer in question and 2. the cost associated with generating the data layer. This information is quite general and can, in practice, vary greatly. However, if it is important for you to do not only your original erosion applications, but also forestry, by looking at those last two rows, you can obtain a general feeling for the amount of effort it will require to be able to support that additional analysis application. (In this case, to do forestry will require the acquisition of a forestry compartment data layer and that will be moderately costly and moderately difficult to accomplish.)

Often a data layer is not necessarily required as an input into an application, but it is easy and inexpensive to generate. Thus, often that layer is created to be used for display and orientation purposes. One may call the installation boundary layer such a map. It may be argued that display and orientation are required for most applications. Thus, a distinction must be made between when layers are used for an application and when they are used for display.

To define what an application is in a GRASS evaluation, let us make the distinction upon reflecting over what makes a GIS valuable. The value of a GIS is to generate new types of information (i.e. an analysis map) from existing information (e.g. from the combination of the soils, slope, and vegetative cover maps). This result existed nowhere else previously. It is new information. A dollar value can possibly be assigned to such an analysis by assigning values relating to the considerations of cost avoidance, increased maintenance management effectiveness or to savings realized by doing the analysis in-house rather than through a commercial contract with its

added overhead costs. In contrast, though it is important, a layer which orients the viewer does not generate new information not available previously. Also the assignment of cost values from display or orientation is not straight forward. In addition, by adopting the strict definition (generating new information) some distinction between the characteristics of a GIS versus a CADD (Computer Aided Design and Drafting) system can be made. The primary purpose of a CADD is to display information visually. The primary purpose of a GIS is to generate new information, which coincidentally is also highly visual in nature. (Clearly this is a superficial distinction. Also, both clearly overlap each other's capabilities even as defined here. But this distinction has a conceptual basis in fact which relates to other questions e.g. most GISs have topologically ferenced data structures while a CADD may or may not).

Thus defined here, an application is an evaluation for a specific purpose which creates new information and for which some cost effects can be identified.

CERL surveyed many projects done over the last several years. Our most completely developed data bases and our broadest applications experiences have occurred at six Army installations and one civil works study area. They are: Forts Hood, TX, Drum, NY, Carson, CO, Riley, KS, Pinon Canyon Training Area, CO, Hohenfelds Training Area, (Federal Republic of West Germany) and Kiethsburg Study Area, IL. These are the locations examined in this paper.

For each location, a survey was done of how each existing data layer in that installation's current data base was used. Table 2 is an example of the information collected for Pinon Canyon. This information was then congregated into application types (Table 3) so that the applications can be compared between different installations.

From this, data layer types can be compared across installations as presented in Table 4, which is similar in layout to Table 1 (with applications versus data types). The difference is that these data types (data layers which have been grouped into data types) already exist as part of some data base and these applications have already been carried out. Thus, Table 1 is a generalized idea of what can be done. Table 4 shows what has actually occurred.

The intersections in Table 4 show the initials of the installations. They indicate which data layers were used for which applications. Multiple labels in a single intersection indicate that usage occurred more than once, either at the same installation or at different locations.

The bottom row of Table 4, the Frequency of data type usage, is a simple summation of the data layer occurrence in that column. This information can be ordered per Table 5 to show the degree of the data type versatility (or value). Table 5 indicates that there are some very highly used data types (e.g. soils data, slope, imagery) and others that seem to be less versatile. (Note that just because they are less versatile, does not mean that they are not important: a wildlife application might not be reasonable without habitats identified.)

There are several different ways of looking at this data and interpreting its meaning. For example, when a person generates a road map and a boundary map, he is really interested in having land use information. Thus the fact that the boundary map had low usage may be misleading. To deal with this question, five general (not necessarily mutually exclusive) groupings were developed: land usage, environmental, natural configuration, topography, and data management. The data types were congregated into these groupings in two ways: 1. under a loose definition of what should be included in that grouping, and 2. under a strict definition of group membership. (i.e. The loose definition is inclusive, the strict is exclusive.) The result is presented in Table 6. The summation of the frequencies of each grouping for either the loose or strict definition suggests that each group has about the same degree of usefulness (i.e. that the summation numbers are roughly about the same for each). The conclusion from this is that a versatile GIS data base needs to include a variety of data layer types.

Conclusions and Recommendations

The value of a GIS is to generate new types of information (i.e. an analysis map) from the existing information (e.g. an erosion potential map from the combination of soils, slope, and vegetative cover maps). Such a result existed nowhere else previously.

Though the display of information is valuable in itself, a sharp distinction was made in this paper between using data layers to display information and using data layers to support an application which generates new information.

Correctly choosing and implementing the data layers will provide greater versatility to pursue GRASS applications. Types of data which have been used at CERL for generation of new data outputs fall into five major groupings (Table 6) each of about equal potential value depending on your location's particular needs. These groups consist of various data types which appear again and again in different applications at many locations. The actual data layer developed depends on the location and the intended applications to which the system will be put. Data layers which are implemented can have other potential usages (Table 1). These other uses will enhance your data base if carefully developed and allow it to have greater usefulness than originally might have been thought possible.

Finally, variations in the data layers adopted will clearly depend on data availability. Though a data layer may be highly useful, if it is not available, alternative configurations will have to be identified. If this is the case, some of the tables presented can be useful to determine how closely related specific data layers are and how alternatives not previously contemplated might yield greater versatility in the data base's ultimate usage.

Reference

Data Availability to Support a Standardized Military Geographical Information system Database, R. Lozar, D Smead, CERL Technical Report N-147, March 1983.

TABLE I - POTENTIAL APPLICATIONS FOR DIFFERENT DATA LAYERS

APPLICATION / DATE LAYERS	Topo Elevation	Topo Aspect	Topo. Slope	Soils	Soils Derived info.	Streams	Wetlands	Open Bodies of Water	Landcover now	Landcover future	Forestry Compartments	Geology	Archeological Sites	Cult. Res. (on Inst.)	Inst. Boundaries	Buildings	Roads	Trails	Firing Points	Range Safty Fans	Range Patterns	Training Compartments	Maneuver Areas	Noise	Outleases	Cult. REs. (off Inst.)	Political Boundaries	Utilities Easements
Archeology	X	X	X			●	●	●				X	●															
Erosion Control			●	●	●				●								●	●										
Forestry	X	X	X	●		X			X		●						X	X										
Change Detection									●	●																		
Fire/Disease Spread	X	X	X			●	X	X	●	X	X			X	X		●	X										
Dredge Disposal	X			X	X	●	●	X	●								X	X							X	X	X	●
Training	X		X	X		●	X		●				X		●		X	X				●	X		X			
Land Condition Trend	X	X	X	X					●	●																		
Noise Impacts														X	●				X					●		●	X	
Facility Siting			X	●	●	X	X		X			X		●	X	X	X											X
Wildlife Habitat	X					●	●	●	●					X											X			
Range Placement	X			X											●			X	X	X	●			X				
Cross Country Move.			X	X	●	●	●	X	●			X	X		X			X					X					
Master Palnning	X		X	●	X	X			●					●	X	X	●							X		X	X	●
County Planning	X	X	X	●	●	X	X	X	●			X	X	X		X	●							X		X	●	●
Visual Impact Ana.	●	X						X	●					X	X	X												
Water Shed Ana.	●	●	●			X		●	X																			
Range Usage Sched.	X		X	●			X		●													●						
Difficulty to generate* in digit form	M	E	E	D	E	M	E	E	M	M	M	M	M	V	E	D	M	D	M	M	M	M	E	M	E	V	E	E
Cost to generate* (*per average)	M	L	L	M	L	M	L	L	M	M	M	M	M	V	L	H	M	M	M	M	L	M	L	L	M	V	L	L

LEGEND:

● required for application

X also needed

Difficulty
E=early
M=medium
D=difficult
V=varies

Cost
Low Med High

Table 2

Example Page of Survey Showing Data Layer Relation to Applications

PINON CANYON

CELL NAME	TITLE	APPLICATION	SOURCE
albedo1	Ground Ref.(LS 10/80)	change detection	LANDSAT
albedo2	Ground Ref.(LS 6/82)	change detection	LANDSAT
aspect	Aspect	background	topographic map
big_arroyo.cl	Class. NHAP Big Arroyo Hills	CERL testing	NHAP
big-arroyo.enh	Enh. NHAP Big Arroyo Hills	CERL testing	NHAP
boundary	Install. Boundary	LCTA	original data
elevation	Elev.-rescaled	background	DEM/DMA
hogback.cl	Class. NHAP(8/83) Hogback	compar. class.	NHAP
hogback.enh	Enh. NHAP(8/83) Hogback	compar. class.	NHAP
landsat.class	LndcovClass. 11/80	background	LANDSAT
lockwood.cl	Class. NHAP 7/83 Lockwood Arroyo	compar. class.	NHAP
lockwood.enh	Enh. NHAP 7/83 Lockwood Arroyo	compar. class.	NHAP
quads	USGS quads	background	original data
ranches	Ranch Houses	training avoidance	original data
range_land	Range/Wdld Sites Las Animas Cty.	potential vegetation	soils.SCS
restrict_areas	Restricted_Areas	land mgt.	training/arch. ecol. study ranch houses
roads	Roads,Trails,Supply Rts.	background	original data
rockcross.cl	Class. NHAP 10/83 Rock Crossing	compar. class.	NHAP
rockcross.enh	Enh. NHAP 10/83 Rock Crossing	compar. class.	NHAP
slope	Slope % derived true elev.	background	topo map
soils.pinon	Pinon Canyon soils-trinad	background	soils.SCS
soils.scs.all	Soils(Las Animas)	LCTA/ background	original data
train.ability	Trnblty.-soils	trainability	soils
train_areas	Training Areas	background	original data
vegetation	derived-NHAP	background	original data
windmills	Windmills	background	original data

TABLE 3 - DATA TYPES USED AT DIFFERENT INSTALLATIONS

Application / Used At	Drum	Riley	Pinon	Kiethsburg	Hood	Hohenfelds	Carson
Noise	X						
Multi Purpose Range		X			X		
Archeological		X			X		
Change Detection			X	X			
Land Condition Trend			X		X	X	
Training/Mobility			X		X	X	
Dredge Analysis				X			
Fanna & Hunting Management			X		X	X	X
Landfill Location					X		
Vegetation/Forestry Management				X	X		
Recreation					X		
Water Management					X	X	
General Data Management			X			X	

TABLE 4 - FREQUENCY OF DATA TYPES USAGE

Application \ Data Types	Noise	Windows	Distance From	Arch Data Results	Install. Boundary	Off-Install. Cultural	Imagery	Install. Land Use	Soils & Reclass Data	Roads	Hydrography	Training Compartments	Dedging	Topo Elev.	Slope	Aspect	Geology	Ecological Sites	Sites Data	Land Forms	Vegetation (Maps	Habitats
Noise	D																					
Multi Purpose Range		R	R	H				R	R	R	R	R			H			H				
Archeological			HR	H		R	R		H		H					HH	H	H				
Change Detection							P												K			
Land Condition Trend					PF		HF		P		F			F	F							
Training/Mobility			H	H				PF	HH P	HH		HH H			HF		H	H			F	
Dredge Analysis			K				K			K	K		K									K
Fanna & Hunting Management								P	C	H					C			H			C	
Landfill Location				H											H			H				
Vegetation/Forestry Management				H					K	H				HK	H			H				
Recreation	H			H											H			H				
Water Management					F		HF		HH	H	HF									H		
General Data Management								P											F			
Frequency of data type usage	2	1	5	7	3	1	7	5	10	7	6	4	1	3	8	2	2	7	2	1	2	1

Legend

H = Hood
D = Drum
P = Pinon
C = Carson
R = Riley
F = Hohenfelds
K = Kiethsburg

Table 5

Data Type Versatility

Order of Data Value

10	Soils (and Reclass)
8	Slope
7	Archeology Data
7	Imagery
7	Roads
7	Ecological Sites Data
6	Hydrography Related and Streams
5	Distance From's
5	Installation Land Use
4	Training Compartments
3	Installations Boundaries
3	Topography
2	Aspect
2	Noise
2	Geology
2	Sites Data
2	Non Satellite Vegetation
1	General Cultural
1	Windows
1	Off Installation Cultural Features
1	Dredge
1	Landforms
1	Habitats

Table 6

Data Versatility Viewed in Different Groupings

GROUP	DATA LAYERS	LOOSE	STRICT
Land Use	Boundary	3	3
	Off Installation Cultural	1	1
	Imagery	7	7
	Installation Land Use	5	5
	Roads	7	0
	Training Compartments	4	4
	Sites Data	2	0
	Windows	1	0
	GROUP TOTAL	30	20
Environmental	Noise	2	2
	Distance From	5	0
	Imagery	7	7
	Aspect	2	0
	Ecological Sites Data	7	7
	Vegetation (Non Satellite)	2	2
	Habitates	1	1
	GROUP TOTAL	26	19
Natural Configuration	Imagery	7	0
	Soils	10	10
	Hydrography	6	6
	Slope	8	0
	Aspect	2	0
	Geology	2	0
	Landforms	1	1
	GROUP TOTAL	36	19
Topography	Soils	10	0
	Topo	3	3
	Slope	8	8
	Aspect	2	2
	Geology	2	0
	Landforms	1	1
	GROUP TOTAL	26	14
Management Data	Archeology	7	7
	Training Compartments	4	4
	Sites-Ecological	7	7
	Sites	2	2
	Habitats	1	1
	GROUP TOTAL	21	21

The Development of the Henrietta Creek Watershed Data Set for use by USDA Soil Conservation Service in GRASS Training.

Richard Franchek

U.S. Department of Agriculture
Soil Conservation Service
501 Felix Street, Bldg 23
P.O. Box 6567
Fort Worth, Texas 76115

ABSTRACT

In the Food Security Act of 1985, the Soil Conservation Service was given the task of identifying all highly erodible soils that were in cropland. Framers and ranchers that fell into these categories are required to have a conservation plan if they wished to participate in any federal farm programs. In October 1987, the SCS started the GRASS Pilot Project with seven States (VT, NY, WA, CO, OK, MO, MI) to identify its use as a GIS tool for helping SCS, State, Area and Field Office personnel with their resource planning requirements. As part of this testing the SCS National Cartographic Center, GIS Staff, developed a demonstration data set for use in GRASS training and as a guide in developing similar data sets for their particular area. This data set was designed with the Field Offices in mind. Data layers were collected to show the uses and potential for GRASS as Field Office resource tool for identifying Highly Erodible Lands (HEL), conservation planning, watershed planning, and information programs.

Since the dust bowl days of the 1930's, the U.S. Department of Agriculture Soil Conservation Service (SCS) has worked with farmers and ranchers in developing and applying conservation practices to prevent excessive soil erosion. SCS field offices have primary responsibility in working with the local soil and water conservation district to implement these practices.

With the Food Security Act of 1985 (FSA) the demand for SCS services has jumped dramatically. The act states generally that to remain eligible for USDA program benefits, a farmer must follow a conservation plan on all highly erodible cropland areas and not drain or convert any wetlands (U.S. Department of Agriculture, 1988). This requires SCS field offices to determine highly erodible areas and to develop the necessary conservation plans with the participating farmer. In some field offices, these determinations exceed 1000 a year. Many are done by hand using existing soil maps and areial photography. The planning process can take several hours. This situation made it necessary for some type of GIS technology that would make the field office planning process more efficient.

In the spring of 1986, the USDA Soil Conservation Service prepared a pilot test plan for evaluating the Geographic Resources Analysis Support System (GRASS) software. In the fall of 1986, the National Cartographic Center (NC) was designated as the GRASS user support center for the seven pilot test sites. These test sites were located in Oklahoma, Colorado, Michigan, New York, Washington, Missouri, and Vermont. Pilot testing officially started in October of 1987 with a

one week training course at the NCC in Fort Worth, Texas. Fifteen people representing the seven test sites attended the training. Te pilot test ended in May of 1988 and final meeting of the test sites was held in June.

The Henrietta Creek Watershed Data Set was developed to support GRASS training as well as provide a resource base. SCS field office personnel could use this data as a guide in developing data layers for their particular area. The watershed covers approximately 17,014 acres in Tarrant County, Texas.

Six polygon layers and two network layers were developed. The polygon layers included soils, landuse, district cooperator boundaries, field boundaries, transportation and watershed area. The network layers were streams an roads. These layers were digitized on ARC/INFO because the MAPDEV portion of GRASS was not yet ported to SCS AT&T 3B2 equipment.

In the development of this data set there was not one single base that was available that would fit all the different source materials. The soils layer was on a rectified photo background (1:20000), the landuse came form color IR high altitude photographs (1:24000), and the rest of the layers from USGS 7.5 quad sheets (1:24000). This condition would be typically encountered by field office personnel when developing their own data layers for GRASS.

The problem was to have all the data layers overlay with each other and still not have to recompile everything to one base. The common theme to many of these data layers was the transportation network. Many of the other layers (district cooperator boundaries, field boundaries, landuse and roads) would have this layer in common when being developed. For example, a particular field boundary would end at the transportation corridor, along with a type of landuse. The transportation corridor would also be a category in the landuse layer. Therefore the transportation network was made a polygon layer and used as a guide for the rest of the intersecting or corresponding layers. As the layers were digitized, all areas that intersected at the road would use the already digitized transportation layer as the stopping point.

When these layers were moved into GRASS, several interpretations were made from the soils layer that included HEL classifications and many different kinds of soil characteristics (such as soil depth, suitability for building sites, range and pasture groups, etc.). With the original eight layers their were now 32 layers.

When training began this data set duplicated real life solutions to problems encountered by people in the field. With the use of masking, windowing, and reporting, HEL determinations were generated using the field boundary layer and the reclassed soils layer. Other types of planning were also duplicated to show the participants that GRASS was capable of helping them with their soil landuse related problems.

Other types of data are scheduled to be included to this data set. Imagery is the next layer for inclusion. This would greatly benefit the user in land cover determinations and for a photo background to conservation planning. also a link to a soil data base is needed to make the interpretations (reclass) of the soils layer more efficient for field office personnel. Finally the data set is to be enlarged to include the portion of the watershed that extends into Denton County, Texas to the north. With this data set the NCC will be able to introduce GRASS technology to the people who will need it the most.

REFERENCES

U.S. Department of Agriculture. 1988. National Food Security Act Manual. Soil Conservation Service. Washington, D.C.

Integration of SPOT Data Into Geographic Information Systems

M. G. Mayers

SPOT Image Corporation
1897 Preston White Drive
Reston, Virginia 22091-4368

ABSTRACT

Low cost and current input data are essential if users of geographic information systems (GIS) intend to support accurate and efficient problem solving. High resolution, current SPOT satellite imagery is provided in standard formats to meet such requirements. SPOT data contains information pertaining to land use/land cover, transportation networks, coastal boundries, urban growth, geology and many other geographic applications. SPOT imagery may be used directly in its raster form for input to a GIS, for example as input to a classification algorithm. SPOT may also be used as a backdrop from which vectors are digitized, such as updating vector transportation network within a GIS environment. The GRASS software package enables a GIS user to effectively exploit SPOT data in both raster and vector formats. The vector and raster capabilities within GRASS enable GIS users to obtain current, powerful solutions to Earth resource problems.

SPOT is an acronym which stands for Satellite Pour l'Observation de la Terre. SPOT is high resolution, land-imaging satellite system designed for Earth resource applications. The SPOT program was initiated in 1977 by the French Government in cooperation with Belgium and Sweden, and is managed by the French Space Agency (CNES). SPOT 1 was launched on February 21, 1986. The spacecraft is operated by CNES which owns the data copyright. The SPOT data distribution is carried out on a commercial basis in the United States by the SPOT Image Corporation located in Reston, Virginia.

SPOT satellite data are available in three formats; digital, film, and print. A single SPOT scene covers 60 km X 60 km swaths on the Earth's surface. The SPOT satellite has two High Resolution Visible (HRV) imaging instruments onboard which provide imagery in two different modes;

1. Panchromatic Imagery 10 meter resolution Single Band Spectral Sensitivity 0.51 um to 0.73 um (Visible Wavelength)

2. Multispectral Imagery 20 meter resolution Three Bands Spectral Sensitivity (Visible Wavelengths)

   Band 1 0.50 um to 0.59 um
   Band 2 0.61 um to 0.68 um
   (Near Infrared Wavelength)
   Band 3 0.79 um to 0.89 um

What are digital SPOT Data? Digital SPOT imagery is stored in raster format. This means a single SPOT image is stored in many small, rectangular, "picture elements" called pixels. Each SPOT pixel represents a reflectance value ranging from 0 (low reflectance) to 255 (high reflectance). When all the pixels are viewed at the same time, they represent a full SPOT image.

Vector format models store data in X, Y coordinates or point, line, and polygon

structures. Vector data models were among the first efforts for producing automated cartographic products (Kolassa 1983, Peuquet, 1984). Today, many geographic information systems utilize vector formats for data storage. Compatibility between different software systems for cartographic products derived from satellite imagery has become a problem (Lauer, 1986). This problem stems form the vast amount of current satellite information available in raster format and map producing software and output devices which utilize data in vector format.

There are many advantages and disadvantages associated with storing and manipulating data in raster or vector format (Kolassa 1983, Peuquet 1984) however, the raster to vector data conversion process typically has two problems associated with it: 1) high processing overhead and 2) difficulty of associating attributes with newly created vector lines, points, or polygons (Kolassa 1983). The best case scenario would be to deal with vector and raster data separately within a single software package. This would allow the user to input, manipulate, output and store digital data in the most efficient format possible, as well as enable co-display of raster and vector files. These raster and vector capabilites are found in the GRASS software package.

Why integrate Raster SPOT Satellite Imagery into a Geographic Information System? There are many justifications for utilizing SPOT Data in a GIS: 1) SPOT satellite imagery is CURRENT. Data may be collected several times per month over the same area by utilizing SPOT's pointable mirrors. 2) SPOT satellite imagery is DIGITAL. Digital data saves time and money by eliminating costly manual processing. 3) SPOT imagery is MULTIPURPOSE. A single SPOT scene may be used for creating land use/land cover maps, extracting transportation routes, monitoring urban growth, planning geological exploration, to name a few applications. 4) SPOT imagery offers GOOD CARTOGRAPHIC ACCURACY. SPOT panchromatic data may be used for mapping at scales of 1:12,000 and smaller. Finally, 5) SPOT satellite imagery is ECONOMICAL. SPOT imagery can cost up to 50% less then aerial photography or acquisition, interpretation, and map generation.

Adavatages of using GRASS to Integrate SPOT Imagery into a Geographic Information System. As mentioned above, there are advantages and disadvantages to storing and manipulating data in raster and vector formats. GRASS has both raster and vector capabilities which enable a user to create new vector files through co-display, create new raster files and update old one files, such as USGS Digital Line Graph (DLG) data. GRASS is being used by a large number of people for a wide variety of applications. This means cooperative efforts in research may be coordinated and some digital data files may be shared. Finally, GRASS is available to the general public at almost no cost, which is ideal for groups operating on a tight budget.

Current, low cost input data are essential if users of geographic information systems intend to support accurate and efficient problem solving. SPOT satellite imagery is current, cost effective, information source for geographic information sytems. Some GIS applications best exploit data in either raster or vector format. The GRASS software package, however, can manipulate data within and among these formats to meet the goals of spatial applications.

References

Kolassa, David A. 1983. The Evolution of Raster Processing Technology within the Cartographic Environment. Auto-Carto 5 Proceedings, Jack Foreman, Editor. Falls Church: ASP and ACSM, pp. 451-460.

Lauer, Donald T. 1986. Applications of Landsat Data and the Data Base Approach. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 8, pp. 1193-1199.

Peuquet, Donna J. 1984. A Conceptual Framework and Comparison of Spatial Data Models. Cartographica. Vol. 21, No. 4, pp. 66-113.

SOFTWARE INTEGRATION

The Design and Implementation of an Exploratory Data Analysis System (S-GEO) for Integrating with GRASS, Remote Sensing and Database Management

Sandra C. Parker

Associate Professor of Industrial Engineering
University of Arkansas
Fayetteville, Arkansas
and
Associate Archeologist
Arkansas Archeological Survey
Fayetteville, Arkansas

ABSTRACT

An exploratory data analysis system, called S-GE, which integrates with the GRASS system, has been designed and implemented. The system is based on the S statistical language developed by Bell Laboratories. S is an interactive environment for data analysis and graphics, and it is it's readily usable graphic capabilities that particularly distinguish it from other statistical systems.

Since GRASS is capable of incorporating remote sensed data, this exploratory data analysis module allows the analysis of such data along with that of a typical geographic data layers, such as soils geology, and other such environmental data. The S module is designed to access the data sets that are transported via the "GRASS to S module" in GRASS. These data sets are in a particular form and S-GEO is designed to recognize and use this unique data representation for information representing the various data layer values for point locations, such as archeological sites, and summary information for all cells of the region of study. Data from GRASS may be transported for any number of different sites lists, each identified by a unique prefix attached to the corresponding data sets.

S-GEO includes various univariate, bivariate, and multivariate analyses modules, all with appropriate graphic representation to aid in analysis, accessible through a menu presentation. Data for different sites list are accessed by identifying the appropriate prefix for the desired data. In addition to access to the geographic data from GRASS, S-GEO also permits the incorporation of other kinds of information regarding sites. This latter information is accessible through a database management system. In this implementation INFORMIX is used to provide the additional data study. The only requirement, however, is a flat file of data with one record per point, so other database management systems could be used. S-GEO includes modules which retrieve and combine geographic data from GRASS with the appropriate descriptive data from the INFORMIX database. One such module utilizes a subset of information, therefore a subset of points or sites, from the descriptive data set and retrieves the appropriate geographic data for these points from the GRASS data structure. These data are then maintained in one structure, the characteristics of which are designed to be like that of data structure resulting from a GRASS to S transport. Therefore the S-GEO modules are moved from GRASS. Another S-GEO module allows the user to identify any particular set of points or sites as a

group to study as a unit. This module retrieves data from both the descriptive data set and the GRASS data to produce a GRASS to S data structure which may also access the S-GEO analysis module.

A geographic information system (GIS) can be defined as an automated system for the storage, retrieval and analysis of spatially referenced data. There are two basic kinds of GISs, a vector-based system and a cell-based system. In the former, areas are represented as vectors (spatial coordinates locating lines) defining polygons, and this type of GIS represented the usual system of choice until recently. The reasons for the early preference of a vector-based system over a cell-based system include the fact that they allow the production of superior cartographic output and they were easy to implement with the hardware and software available prior to the mid-1980s. A major shortcoming of vector-based systems, however, is that they do not allow for appreciable statistical characterization or analysis of the geographic data which they store.

A cell-based GIS is one which effectively divides a spatial surface into a rectangular grid and stores, for each grid cell, the spatial coordinates which locate the cell on the surface as well as any discrete or continuous data pertinent to describing or characterizing the surface of the cell. These data effectively provide an n-dimentional matrix which can be used to produce high quality visual output, given high resolution visual display equipment, and can be used as the basis for extensive statistical characterization and analysis.

During the 1970s and early 1980s, cell-based systems were constrained by factors such as inflexible cell sizes, large data storage requirements, and the high cost of raster (cell) display devices. Though these factors limited their application, cell-based systems were nonetheless considered to have greater analytical potential due to the ease with which the data could be subjected to numerical analyses. This potential was demonstrated by developments in the area of remote sensing, including the application of classification methods to analyze remote sensed scenes.

Hardware developments in the mid-1980s and software written to utilize this hardware, have allowed the emergence of cell-based GISs which address the earlier problems. Modern cell-based systems can use the many classification methods developed in remote sensing as well as a variety of map calculator operations. However, there has been no cell-based GIS development to date which encompasses the full range of robust multivariate statistical techniques which are applicable to the multidimensional GIS data. This paper reports a software development which is designed to provide the full range of statistical techniques to a particular GIS, Geographic Resources Analysis Support System (GRASS 3.0).

GRASS is essentially a cell-based GIS which is interactive and allows for the visual presentation of grid cell data layers in two dimensions, as well as 3-dimentional visual presentations. Vector and point data may also be stored and displayed. Numerous map layer types may be represented for an area of interest; some data layers represent basic raw data collected from the surface, such as elevation data, while other data layers are derived from one or more other surface representations. Examples of derived data layers are slope and aspect, both derived from the basic elevation data. Additional map layers may be derived by classifying cells according to their distance to some aspect of the landscape such as streams. An analyst may also reclassify a data layer, such as soils for example, into a new data layer which gives a different organization of the basic information. One might derive a more general soils map which groups similar soils into fewer classes, but classes which may have more interpretive meaning in the analysis than did the more finely divided original soils map.

In most GIS applications, there are many data layers of interest, particularly given the ability to derive new layers of analytical interest. The resulting complexity

and multidimensionality of GIS data sets can make multivariate patterning difficult to assess. Modern statistical developments in exploratory data analysis (EDA) include techniques which can be useful in analyzing multivariate spatial data. In fact, there is an EDA system which shares some of the characteristics of GRASS - its interactive and analytical attributes. This EDA system is the S system (Becker & Chambers 1984), developed at Bell Laboratories, and it is interactive and strongly graphics oriented. The linkage of a powerful cell-based GIS and an interactive, graphic statistical system such as S is both needed and feasible. A GIS should not be seen as an isolated system but rather as one component in a comprehensive automated database environment including database management, remote sensing, and EDA components. Toward this end, software to link GRASS, a database management system, and S has been developed and implemented by the author.

An exploratory data analysis module, based on the S statistical language, has been developed to link S procedures to GRASS databases as well as to other databases which pertain to points of interest located in the analysis surface. Since GRASS is capable of incorporating remote sensed data, this EDA module allows the analysis of such data along with that of typical geographic data layers, such as soils, geology, and other such environmental data.

The new software is designed to access the data sets that are transported via the "GRASS to S module" in GRASS. These data sets are in a particular form, and the new EDA module is designed to recognize and use this unique data representation for information representing the various data layer values for point locations, such as archeological sites, and summary information for all cells of the region of study. Data from GRASS may be transported for any number of different sites lists, each identified by a unique prefix attached to the corresponding data sets. The user is queried as to the prefix or prefixes desired for any particular analysis, and this is the mechanism for control of the content of the data matrix to be subjected to analysis. The new EDA software includes various univariate, bivariate, and multivariate analyses macros, all with appropriate graphic representation to aid in analysis, accessible through a menu presentation.

In addition to access to the geographic data from GRASS, the EDA module also permits the incorporation of other kinds of information regarding sites. This latter information is accessible through a database management system. In this implementation INFORMIX is used to provide the additional data regarding the points of interest in the geographic region under study. The only requirement, however, is a flat file of data with one record per point, so other database management systems could be used. The new software includes macros which retrieve and combine geographic data from GRASS with the appropriate descriptive data from the INFORMIX database. One such macro utilizes a subset of information, therefore a subset of points or sites, from the descriptive data set and retrieves the appropriate geographic data for these points from the GRASS data structure. These data are then maintained in one prefixed structure, the characteristics of which are designed to be like that of the data structure resulting from a GRASS to S transport. Therefore the EDA macros are equally usable with such a data set as with the data that are moved from GRASS. Another macro allows the user to identify any particular set of points or sites as a group to study as a unit. This macro retrieves data from both the descriptive data set and the GRASS data to produce a GRASS to S data structure which may also access the EDA analysis macros.

In summary, this EDA module provides a link between a powerful cell-based GIS system with the capability to incorporate remote sensed data, discrete and/or continuous point data maintained in a database management system, and a powerful interactive, graphic EDA system. It provides an analytical database environment from which one can do a wide variety of analyses, including multivariate analysis of GIS data.

REFERENCES

GRASS 3.0, United States Army Corps of Engineers, Construction Engineering Research Laboratory, Champaign, Illinois, 1988.

Becker, R. and J. Chambers 1984 S - An Interactive Environment for Data Analysis and Graphics. Duxbury Press.

THE MARRIAGE OF GRASS AND ORACLE

Sanford Fidell, Michael Harris, and Nicolaas Reddingius

BBN Systems and Technology Corporation
21120 Vanowen Street
Canoga Park, CA 91303

ABSTRACT

GRASS 2.0 was ported to a PC/AT under MS-DOS as part of the development of a prototype Assessment System for Aircraft Noise for the U. S. Air Force. Information about and complementary to map layers are stored in an ORACLE relational database.

The 1969 National Environmental Policy Act (NEPA) requires the U.S. Air Force to assess the effects of their operations on people, animals, and structures. Environmental planners prepare analyses that range from simple findings of no significant impact to preparation of extensive environmental impact documents. The analyses are typically based on combinations of geographically referenced and other documentary information.

Planners at remote Air Force facilities are often hard pressed to obtain the timely, reliable, and comprehensive information needed. They often need to present composite maps displaying information from disparate sources (topography, transportation networks, aircraft noise exposure contours, aviation, population, land use, archeology and cultural features, seasonal habitats of endangered species).

It is not enough to locate and identify potential aircraft noise impacts, one must also evaluate their significance. This requires extensive computation, but also access to various engineering models and reference to citations from specialized technical literature. Few computer-based tools are currently available to facilitate these analyses.

The U.S. Air Force Noise and Sonic Boom Impact Technology Program (NSBIT) is sponsoring development of an Assessment System for Aircraft Noise (ASAN) to provide planners with the means to conduct their environmental assessments in an efficient and cost-effective manner. The first part of the development effort, development of a preliminary prototype system, is the topic of this paper. Development of a final prototype is now underway, in preparation for delivery of the first release of a production system in 1992.

The following were among the fundamental design constraints on ASAN:

- o users may not have extensive training in environmental acoustics and NEPA requirements.
- o users often work on more than one environmental at a time.
- o duration of assessments may exceed the job tenure of the planner.
- o users, often junior officers trained in civil engineering, cannot be expected to have any great familiarity with computing.
- o users cannot be expected to have access to advanced graphic workstation hardware.
- o and, to keep things challenging, we had less than six calendar months to put the prototype together.

The design goal of ASAN is to provide tools to assist planners in their tasks:

1. Process oriented.
   - o Determining the steps required to complete the assessment.
   - o Tracking the status of the work through its stages.
   - o Helping find likely sources of information for unstructured parts of the data gathering and analysis.
   - o Organizing "fixed" data maps, mission profiles, aircraft noise characteristics) effectively.
2. Analytic.
   - o calculating to predict noise exposure based on current best engineering practice.
   - o assessing of impacts based on current knowledge of the effects of noise exposure on the environment.
   - o reporting the engineering and assessment results in a form useful for inclusion into an final document (e.g. an Environmental Impact Statement or Finding of No Significant Impact).
3. Organizational.
   - o documenting work and decisions for review by cognizant authority.
   - o record keeping required for compliance with NEPA.
   - o making supporting information accessible and understandable to successors after a planner has been reassigned to other duties.

From the user's perspective, ASAN's appears as shown in Figure 1.

The ASAN target system was a Zenith Z-248 (USAF's PC/AT "compatible" computer) running Version 3.3 of MS-DOS. This relatively inexpensive machine was selected 1) because it has been purchased in large quantities by the Air Force and is thus likely to be generally available throughout the environmental planning community, and 2) because the commercial mass market for this and other IBM PC/AT-compatible computers makes available low cost graphics, mass storage, and other peripheral equipment.

It was decided to develop ASAN on top of existing GIS and relational database management systems. GRASS 2.0 was selected as the GIS system, ORACLE as the relational database manager, and U, a BBN-proprietary screen-oriented user interface for the user dialogue. ASAN's software architecture is illustrated in Figure 2.

MicroSoft C was chosen for the implementation under MS-DOS, largely because of its completeness and close compatibility with standard UNIX C. In the past we have ported software written in C for UNIX to the MS-DOS/MicroSoft C environment without substantive change.

The original battle plan for integrating GRASS into ASAN was to port as much as possible of the GRASS software from its original UNIX/Masscomp environment to the Zenith Z-248 under MS-DOS. Then, as analytic programs were developed and needs for geodata manipulation identified, software interfaces could be built to the appropriate GRASS features. A major advantage of this plan was the availability of a stand-alone GRASS system on the ASAN host during development. This could provide views of ASAN data that were not directly available through ASAN itself.

Like many battle plans, this one did not survive long after initial contact with the enemy, for several reasons. GRASS was apparently developed as a set of independent functions. Users interact with a consistent top-level shell or control interface that selects and invokes functions. The control interfaces inside most functions are not standardized, however. Some functions take LISP-like statements as user input, others take command lines of the keyword-and-arguments form, still others use their own menu structures.

Worst of all from the perspective of building ASAN on top of GRASS, control interfaces for many functions are deeply embedded in the functions' execution logic. In some cases, modifying the original code to provide compatibility with the ASAN standard user interface proved to be more costly than re-implementing the functionality from scratch.

GRASS 2.0 code also reflects several assumptions concerning the supporting

hardware and software environment that are inconvenient for porting efforts. Parts of GRASS rely on a UNIX interprocess communication facility not available in MS-DOS, mostly to support a multi-user environment with a single shared display device. In the single- user ASAN system, this and several similar issues were easily resolved. Dependencies on idiosyncratic Masscomp graphics protocols were expected, but surprisingly few were found. Those that were found were not of a serious nature.

Much more serious was the treatment of pointers and integers as interchangeable and equivalent quantities. This practice is forbidden by the C language definition but is not detected by most C compilers. This assumption rarely causes problems on 32-bit computers, where both pointers and integers are 32 bits wide. But in our development environment, it created some of the nastiest porting problems. Undetected instances of interchanging pointers and integers can cause either constant or intermittent unpredictable operation, inconsistent program output, data corruption, damage to unrelated co-resident software, system software crashes, and even disk and firmware damage.

More than two hundred instances of this problem were found in the GRASS source code. Many instances were found by running the UNIX Lint utility over the source code and carefully investigating every reported anomaly. Critical sections of code were analyzed line by line by an experienced programmer.

The GRASS code is structured as a multilayered model. An action that produces graphics can pass through many layers of function calls before drawing commands are produced. In the interests of simplifying the MS-DOS version, it was decided to reduce the number of functional layers as much as possible. One obvious reduction was the combination of the drawing-command layer on the user side of the graphics pipeline with the actual drawing code on the device support side. After resolving considerable confusion over the term "window" (GRASS uses the term "window" to refer to the entity called a "viewport" in classical computer graphics), the GRASS window layer was essentially eliminated.

The usual difficulties associated with porting a large UNIX-based software system to MS-DOS were also encountered. The most notable were in the area of memory management. The Microsoft C implementation of the standard C memory management functions -- malloc(), calloc(), free() -- well known to be unreliable in large complex software packages, was indeed troublesome in this port as well. Since the problem is related to the MS-DOS intrinsic memory limit structure of 640k bytes in 64k segments, it can be thought of as a memory utilization problem.

One solution would be to add expansion memory hardware to the system, and to use a memory management software package that utilizes the expansion memory for dynamic allocation. Unfortunately no such package is available, and neither time nor funding was available in this project to develop one. The recent release of MS-DOS version 4.0 may present another solution, since it reportedly includes an expansion memory handler. But for the ASAN prototype development effort, the simplest solution was adopted: problematic memory management code was rewritten to employ static memory.

Recall that the goal of this effort was to construct an ASAN concept demonstration prototype system in a short time with minimal resources, and not specifically to port GRASS to the MS-DOS environment. As more and more porting problems were encountered, the development process was necessarily refocused on the overall project goal, and the GRASS presence in ASAN was drastically reduced, until the goal of developing an independent GRASS within ASAN was abandoned. If a GRASS feature was not required for the demonstration, it was omitted.

In the case of absolutely essential GRASS functions, such as map layer display management, it eventually proved to be quicker to write new code from scratch rather than to unravel and port the original GRASS code. As a result, the ASAN concept demonstration prototype system software does not contain any original GRASS code, even though GRASS geodata

file structure concepts and file formats were retained. Raster data are stored as "cell" files; vector and point data are stored as "digit" files (used by standard GRASS as an intermediate format in some internal and conversion processes). In this manner, the data bases developed for the ASAN prototype are not lessened in value, and in fact can be further enhanced on standard GRASS workstations.

As useful and effective as a GIS is for many ASAN-related purposes, its capabilities fall short of an environmental planner's needs. For example, much of the information is not inherently georeferenced. e.g., aircraft noise and performance characteristics. ASAN also contains an elaborate citation index to the technical literature, a Rolodex-like capability, and various other non-georeferenced information. A conventional database is more appropriate for such information than a GIS.

Some other limitations of geoinformation systems in general and GRASS in particular also limit the usefulness of a purely GIS approach to solving the environmental planners' problems. For example, the range of sound pressures of interest can vary over a range of about 15 orders of magnitude, while small local variations are important. This precision is beyond the capacity of the GRASS cell file structure.

Further, analyses of noise impacts on buildings, ruins, or historical sites often require consideration of a large number of parameters. For example, great detail might be required about the nature of a structure; its age, construction, and use; its occupants and times of occupancy; and so forth. In addition, the sources of information and dates of entry or modification must be tracked for several purposes: to create a complete record of decision, to evaluate currency, and to facilitate performance monitoring and updating. It is convenient to manipulate this mass of detail in ways that cannot be readily accommodated within a GIS.

Non-georeferenced information in ASAN is stored in an ORACLE relational data base, which provides a full implementation of the ANSI standard Structured Query Language on a PC. A unique feature of ORACLE is that most of its activities take place in extended memory, running in protected mode on the 80286 or 80386 microprocessor. In other words, except for 70 kB or so, the MS-DOS operating system does not "see" anything of the database software.

This was a major consideration in ASAN development, because the SQL engine is so large that it precludes development of large scale applications that must share memory with the DBMS. Since the aim was to build a large simulation model on top of the DBMS (not merely to use a 4GL to do simple stores and retrieves), this proved to be an essential feature of the database system.

The GRASS-ASAN-ORACLE prototype stores non-georeferenced data in normalized relational tables, where it is accessible to a large number of analytical routines. ASAN uses the relational approach for discrete geo-referenced data, but creates a GRASS file as needed for display. A structure similar to the cell file, capable of storing floating double rather than integer data, is ultimately required for our continuous geo-referenced data.

While a normalized relational database possesses a certain theoretical cleanness, this comes at the price of system overhead. For a demonstration prototype this is not necessarily a problem, since such a system requires only limited amounts of data, and an easily modifiable database is more important than high speed performance.

Our approach violates nonetheless some basic principles of database design by carrying large amounts of redundant information in GRASS and ORACLE files. A set of ORACLE tables is therefore required in ASAN to maintain synchronization within the redundant portions of the database. This is not an undue burden because ASAN must keep track of revision levels, entry dates, and the "freshness" of the data in any event. ASAN continually checks that one piece of information does not conflict with another, and that a particular set of user-specified information is in fact a valid combination of possible inputs. An additional check for database redundancy fits naturally with this requirement.

Of course there is one major caveat: One can view information directly with either the geodatabase or conventional database systems, but it is an absolute necessity that the ASAN shell be used to modify information. The internal audit trail loses file synchrony if ASAN is bypassed to update them.

Storage of information about the various map layers in ORACLE permits formulation of interesting questions from the GRASS/ORACLE combination. For example, consider a map layer containing the habitat of an animal species. When that map layer was created it was recorded in the ORACLE database that it deals with a particular animal or group of animals. (This is done by including the animal's taxon number from ASAN's taxonomy table for the animal kingdom.)

ASAN can not only answer the question "Where does this animal live?", but can also retrieve references in the scientific literature concerning the animal from a bibliography that is indexed by taxon number. The production version of ASAN will be able to carry this process one step further. ASAN will eventually be able to answer the question "What do I need to read to understand the impacts on the local fauna?"

The preliminary prototype was first demonstrated last winter. It performed all of its intended functions and was received with great interest by its sponsors and members of the environmental planning community. The design of a first operational prototype is now under way.

The Air Force is in the process of standardizing on 80386 machines. This hardware is a much better engine for large and compute-bound systems such as ASAN. MS-DOS, with its 640kB limitation, does not make much sense in this environment, but the Air Force has not yet committed to an operating system for the new machines. Development of ASAN will probably not occur within OS/2, which is at present only partly developed itself, and may not be the system of choice in 1992.

Continuing development of ASAN under MS-DOS imposes the operating system's limitations on ASAN's architecture. Although it is preferable to design a software- intensive system for delivery in 1992 for the 1990's desktop machine, development cannot continue in an operating system vacuum. One interim solution to this dilemma is to conduct the next development phase of ASAN in UNIX. This would allow ASAN to sit out the battle in the operating systems market for as long as possible. It also means less work porting new GRASS developments to the eventual operating environ-ment.

Since GRASS is not being utilized within ASAN as a stand-alone system, but rather as a building block for a much larger system with its own user interaction, development of a "layered" GRASS (in which function control logic is separate from execution logic) is highly desirable. A suggested model for future development of GRASS is the approach taken by conventional databases. A modern DBMS has a 4GL front end so that functional primitives can also be called from application code through a well- defined protocol. A GRASS that includes user access to its component building blocks would be a major step toward bringing geo-information systems to a new level of functional maturity and making GRASS accessible to many other problem-specific applications.

ACKNOWLEDGEMENT

The effort reported herein was sponsored under Air Force Contract F33615-86-C-0530, administered by the Noise and Sonic Boom Impact Technology Project Office at Wright- Patterson AFB, Ohio under the direction of Lt. Col. Geral Long, and technical contract manager, Mr. Lawrence Finegold.

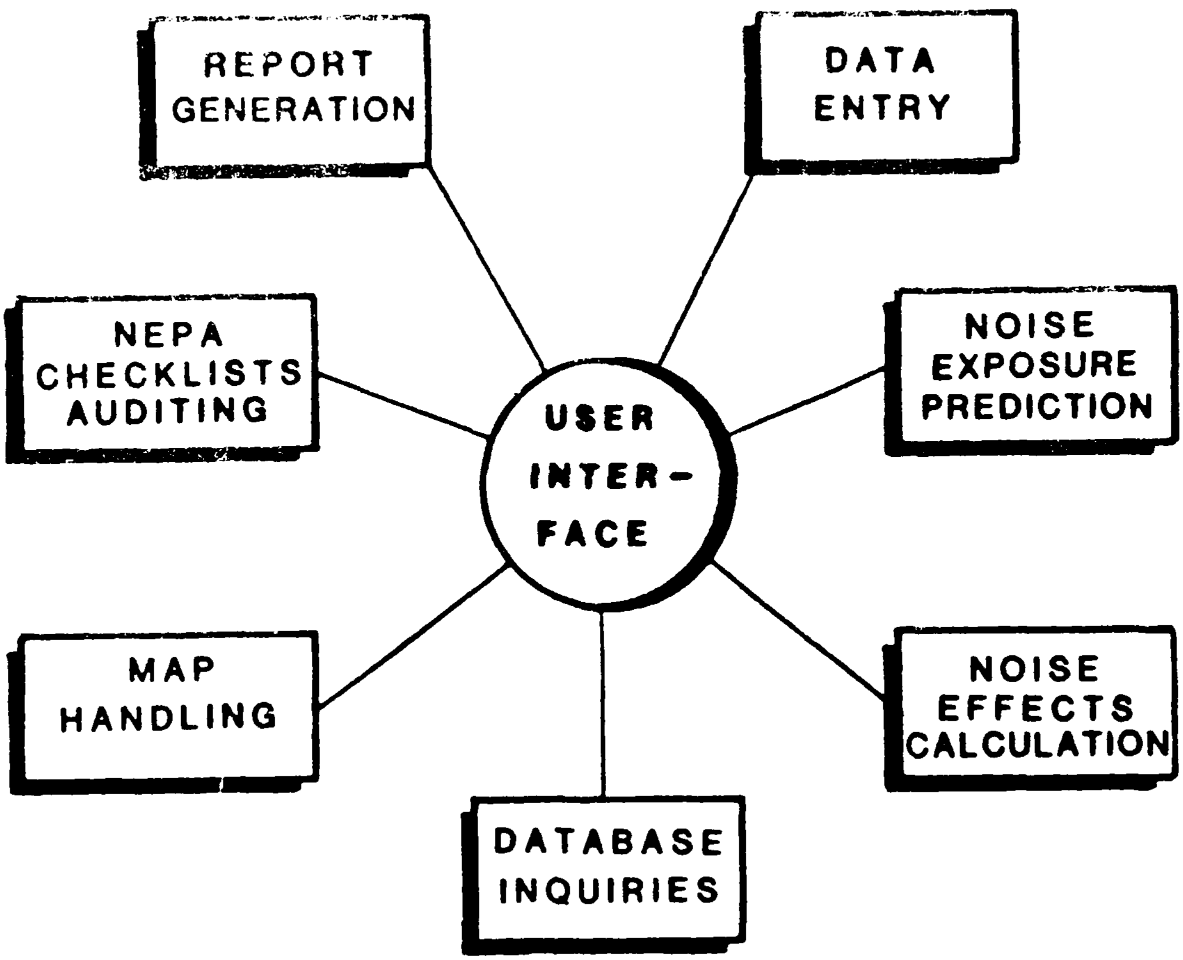

Figure 1: ASAN from the User's Perpective

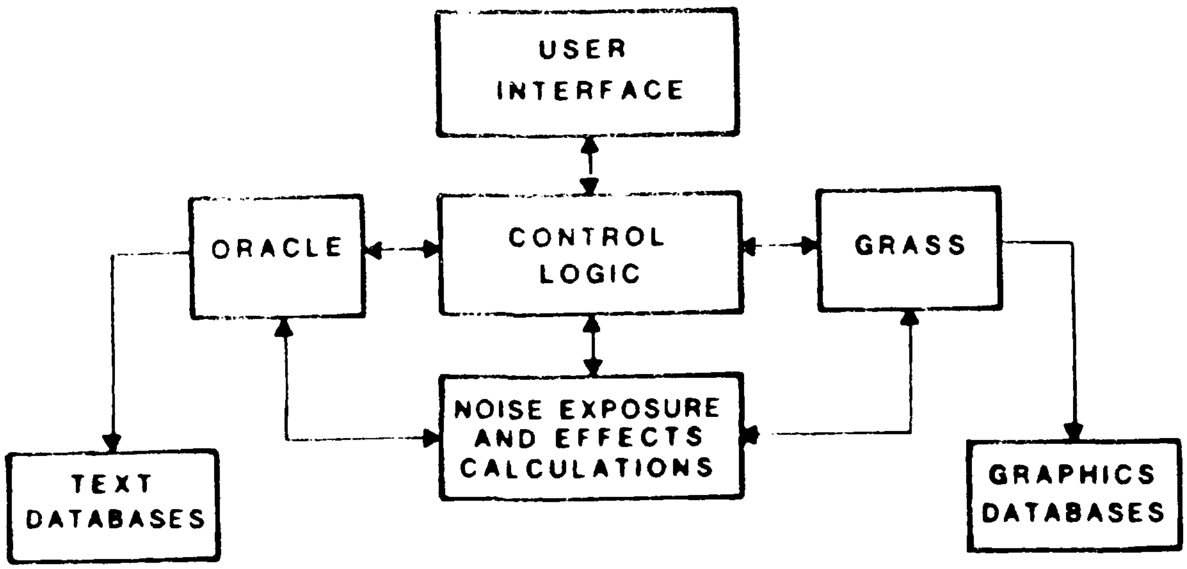

Figure 2: ASAN Software Architecture

INTEGRATING RELATIONAL DATABASE CAPABILITIES WITH THE GRASS GEOGRAPHIC INFORMATION MANAGEMENT SYSTEM

James A. Farley

Arkansas Archeological Survey
Fayetteville, Arkansas

ABSTRACT

Relational database management systems (RDBMS), have provided users with the ability to store, manage and manipulate large volumes of information for nearly twenty years. Since E. F. Codd's early research with the relational model in the late 1960's these systems have been refined and today represent a standard for information management. While RDBMS are traditionally viewed as text based systems they can contain spatial reference data which provide a logical link to a GIS such as GRASS. This ability to link text based systems with a GIS serves to extend the range of information available to the GIS user. For example, UTM registered point data and polygon or area data can have multiple attribute associations which are retained in the RDBMS and accessed from within the GIS.

The Arkansas Archeological Survey is currently developing an interface between GRASS and a number of archeological databases constructed using the INFORMIX relational database software which supports the Structured Query Language (SQL) standard. This interface will allow the GRASS user to generate sites lists and reports using SQL from within GRASS. Discussion will focus on the rationale for this approach, the benefits it will bring to GRASS user and the progress made to date.

INTRODUCTION

The Arkansas Archeological Survey is in the final stage of designing an extension to the GRASS GIS which will provide an interface to a number of archeological databases. These databases have been developed using INFORMIX, a commercial database package distributed by INFORMIX Inc., of Menlo Park California. The INFORMIX software supports the Structured Query Language standard (SQL), and adheres to the relation model popularized by E. F. Codd in the late 1960's. This paper will examine both the rational behind the development of an extension to GRASS which supports an interface of this type and the methods and strategies we employed in designing this interface. Although the discussion will focus on archeological applications using particular databases, the need to link a relational database to GRASS is by no means peculiar to archeology. In fact the interface presented here will be generalized to accommodate many diverse applications. There are three key issues which should be should be kept in mind throughout the discussion:

1) GIS applications may be appropriately applied to any problem or data set which has a spatial component regardless of scale.

2) A GIS is nothing more than a spatially registered database.

3) The cost associated with constructing a database, spatial or non-spatial, demands that the information be used in as many contexts as possible.

Underlying these issues is a general philosophy which emphasizes a modular approach to information management and automated analysis. Within this conceptual framework two major themes are relevant to the present discussion. First, any database represents a long term institutional investment and because of this it must be maintained in a manner that permits maximum usage in a wide variety of applications. This can be thought of as a corporate strategy for database management. Second, in order to facilitate such a strategy a toolbox mentality needs to be developed in which individual application programs are seen as separate but compatible tools which can be used in an integrated fashion to accomplish research or management objectives. Under this format each tool (the GIS, the RDBMS or other applications programs), is used in a way that emphasizes its strength and no tool is modified in order to perform tasks for which it was not originally designed. This is the inverse corollary to the K-TELL GINZU it slices it dices paradigm of inertia. Simply stated, when you need a hammer use one, but when you need a screwdriver don't try to make the hammer do the screwdrivers job.

A DATABASE INTERFACE TO GRASS, WHO NEEDS IT?

GRASS is an extremely powerful and comprehensive software package which is capable of manipulating spatial data to address a wide range of problems. Given this, why introduce additional features which add to its complexity and may be responsible for making the system more difficult to use? This approach can be characterized as the "IF IT AIN'T BROKE ...DON'T FIX IT paradigm of progress", which has a great deal of merit particularly where computers are involved.

However, there are several valid reasons for such an undertaking. First, the concept of a text based or relational database is almost synonymous with computer applications. Database's, particularly relational databases have become an integral element in the day to day operation of businesses, the Federal government and research outlets across the country. In these wide ranging contexts people have amassed an extraordinary volume of data and have come to rely on databases to maintain multiple observations on diverse data relationships in an organized structure. This organization of information is the job that database tools are carefully crafted to perform. If we as GIS proponents were to adopt a corporate attitude towards both the spatial and non-spatial databases we have access to we might derive significant benefits from our attempts to exchange data and information between the two.

In addition to these basic organizational strengths relational databases often contain information which is closely linked to GIS applications. For instance, much of the information stored in databases is spatially registered or has a readily identifiable spatial component. This is particularly true in the area of natural resource management. Archeology, for instance, attempts to register everything to a particular provenience and we have entered vast quantities of data dealing with individual artifacts, subsistence strategies, settlement patterns and population dynamics into relational databases.

Each of these entries has corresponding locational information and while the scale may vary from centimeters to meters to kilometers the simple presence of this spatially registered data demands that the information be subjected to the type of spatial analysis normally associated with GIS applications. It is suggested here that this preponderance of locational data located in text based databases is not unique to archeology, but might be widely found in any number of applications oriented towards the environmental sciences.

Most importantly there is a logical linkage between databases and GRASS which is evidenced by some of the existing GRASS tools. The SITES module requires that a series of UTM coordinates with an accompanying tag or identifier be entered into the system to prepare a surface suitable for positional analysis. The RECLASS module requires that a series of relationships are defined and input to modify existing data

and construct a new map layer. Each of these processes involve the input of information which is maintained outside of the GIS. In many instances this information has been previously entered into an existing non spatial database of one type or another.

While GRASS is perfectly capable of handling these situations in its present configuration the methods used to conduct a RECLASS or to generate a SITES list is somewhat analogous to the hammer and screwdriver metaphor cited above. The strength of GRASS lies in its ability to manipulate and display spatial data. The strengths of a database lie in its ability to perform conditional subsets, to adjust values based on selected criteria and to quickly generate reports with multiple formatting options. These features are exactly what is required by both SITES and RECLASS. Therefore establishing a link to a text based database is a logical next step in the evolution of the GRASS system. Taking this step should improve the ease with which modules such as SITES and RECLASS are executed and in addition should stimulate increased use of these tools while incorporating a much wider range of valuable data. This type of symbiotic integration is an excellent example of the TOOLBOX approach to automated analysis.

DBGIS TOOLS: AN INTERFACE ONLY A MOTHER COULD LOVE

Integration is a lot like time, it's a relative phenomena. Two things can be characterized as integrated if a simple connection is made between them. However, this is not the degree of integration we felt would be most likely to encourage use of a database link with GRASS. GRASS users are already accustomed to an extremely flexible, friendly interface which accommodates many levels of user sophistication. Because of this we felt it was essential to retain as much of a GRASS feel as possible leaving the actual database interface transparent to the end user where ever feasible.

The resulting suite of DataBase - GIS tools (DB tools), attempts to achieve this level of integration while addressing some of the issues outlined above. These tools are designed to be accessed from within existing GRASS modules or from new modules which simulate those already present in GRASS. For the purposes of development we have relied on two existing databases which have been used in ongoing research at the Arkansas Archeological Survey. The AMASDA database is a comprehensive inventory of over 22,000 archeological sites with information measuring over 130 attributes including UTM easting and northing. The second database used during development is smaller and contains census oriented data from over 700 counties encompassed within the COE Southwestern Division. In this case each county polygon received a unique identifier enabling us to treat this data in much the same way as one would treat a map layer representing training areas on a military installation. However, as noted above the usefulness of these tools is by no means restricted to such a narrow range of application and they are being implemented with an eye towards eventual generalization.

DB SITES - Database Derived SITE LISTS

The DB SITES module will be created as an extension to the existing SITES programs appearing as option number nine on the menu. This program will be driven by a direct interface to INFORMIX and will support a query by forms format. Under this format the GRASS user will be temporarily placed in the INFORMIX environment and a data entry form will appear on the screen. At this time the user will select the database tables from which to draw information and enter values into the fields upon which the query is to be based. In most instances this process will require only a few key strokes. The database system will then take control and retrieve the rows which satisfy the query and exist within the currently defined window. For example in AMASDA one might request all sites with mounds or all sites that have been assigned to the cultural period Early Caddo. More complex queries might seek all sites which have truncated mounds, and have been designated as Late Caddo and also have

been subjected to nonscientific investigations. Such a query would allow us to rapidly identify all known Late Caddo mound sites which have been the focus of non-authorized excavations. The query itself may incorporate wildcard characters, range values, character strings or specify Boolean relationships. Once retrieved the rows will be formatted to conform to a sites list structure using the INFORMIX report generator ACE and placed in the site_lists directory of the current LOCATION - MAPSET. At this time the user is returned to the main SITES menu and may begin to work with the newly created sites list.

Although this module has not been formally put in place we have used its basic compo- nents in conjunction with AMASDA. The results of this limited testing were very encouraging. Over the course of an afternoon we were able to generate approximately thirty separate sites lists using both simple and complex queries to the database. I want to emphasize that these sites lists were generated using only the most basic types of queries and that they have been developed for the purpose of example. We have not yet begun to truly exercise the database to take advantage of the multi table querying capabilities it supports to extract complex associations from across the state.

DB RECLASS - Dat base Derived RECLASS of Polygonal Data

The DB RECLASS module will automate the existing GRASS modules Reclass (G_reclass), and will permit the user to specify any associations which exist in the database as the criteria for creating the new map layer. This process will simulate a GIS which preserves multiple attribute tags for individual polygons.

DB RECLASS will be accessed from the command line using an argument specifying the GRASS map layer which is to be the object of the reclass operation. Like DB SITES, DB RECLASS will place the user into the INFORMIX environment. At this time the user may enter a single value, multiple values or a formula which is to form the basis of the reclass operation. This information is then incorporated into an SQL statement and submitted to the database which performs the requested operations and transfers control to the ACE report writer which formats a Greclass input file. Greclass is then executed using this file as input and upon completion all necessary GRASS support files are constructed and control is passed back to the command line.

Using this approach in conjunction with the COE Southwestern Division census database we were able to quickly construct a number of choropleth maps which represented demographic change measured at ten and twenty year intervals and a number of archeological maps depicting site densities and levels of investigation throughout an eight state region. Because each polygon, counties in this instance, has a unique number and associated attribute information the reclass operation may be quickly performed by the database along a number of important dimensions.

Again, this approach is not restricted to this narrow range of application. We have begun to experiment with this process using data derived from soils maps and county soils association records and have been successful in quickly creating map layers depicting phenomena such as Ph and soil texture. Because DB RECLASS will accept an algorithm which is used to derive new values to represent each polygon it will function in much the same way as Gmapcalc except that the calculations are performed on columns within the database instead of between map layers in the GIS. For example in order to standardize the census data for area we had the database divide the population value for each county by the number of square miles within that county to obtain an estimate of population per square mile.

DB WHAT - Mouse Driven Database Attribute Reports

The final database tool we are working towards at this time is a variant of the Dwhat Command supported by GRASS. The DB WHAT module will function in much the same way as Dwhat except that it will poll the database and report back on

any known attribute associations within fixed window of one kilometer. Like Dwhat this will be a mouse driven routine. The user will enter the DB WHAT command from the command line and a graphics cursor will appear on the display screen. The user then positions the cursor in the approximate area for which information is wanted and clicks on the mouse. At this time the UTM coordinates associated with the graphics cursor position are passed to INFORMIX and SQL statements are executed to obtain all the rows which fall within one kilometer of that location. All columns for the resulting rows are then formatted using the ACE report writer and output. Ideally this module would report directly to the screen, however differential processing speeds and the potential for a wide range of variation in database size may require the database search and report generation to run in the background and write to a disk file.

We feel the DB WHAT module will provide a great deal of insight into the range of information available in the vicinity of any given location within the GIS. It will function as a snapshot device which allows the user to simultaneously examine a GIS map surface and a suite of associated attributes which are stored off-line in the relational database.

INTERFACE ERRATA - SO MANY TOOLS SO LITTLE TIME

The tools discussed above are only a small fraction of those which could, and hopefully will in the future, extend the capabilities of the GRASS system. In fact, the interface we are developing might most effectively be viewed as a proof of concept. The concept in this case is not just the linkage of GRASS to a relational database, but rather the idea that the integration of a number of specialized tools can enhance the environment for information management and automated analysis. We have also been working with linkages to the "S" exploratory data analysis system and have discussed the possibility of working towards an interface to a desktop publishing system. With the arrival of X-windows it is conceivable that a windowing environment could be established which allows the user to work at a single display which is running GRASS, a relational database system, an exploratory data analysis package and a desktop publishing system. In such an environment the concept of integrated tools might realize its full potential permitting multiple applications programs to execute simultaneously while data is input and output from one application to another.

Imagine the possibilities!

GRASS in the X-Windows Environment Distributing GIS Data and Technology

Kenneth Gardels

Center for Environmental Design Research
University of California
Berkeley, California 94720

ABSTRACT

GRASS (Geographic Resource Analysis Support System) is an interactive, high-performance Geographic Information System package designed to operate on a variety of based graphics workstations. Recently, work has been under way at UC Berkeley's Center for Environmental Design Research to migrate GRASS to the *X-Windows* environment, which is emerging as the standard graphics, windowing, and user interface system for workstations. *X-windows* provides a number of advantages to the GIS application developer and user: (1) applications software developed for one hardware platform is portable virtually without modification to any other platform running *X*; (2) *X* is a networked-based protocol, meaning that users can directly access data and software residing at any remote host and display results locally; and (3) a consistent user interface may be maintained across a range of platforms and implementation schemes. This paper x will discuss the issues in moving GRASS to the *X* environment, both in terms of the graphics requirements and the development of a user interface. It will then describe a series of implementation scenaria for GRASS, including remote database query and geoprocessing with workstation-based interaction, GIS database distribution for local processing and display, and local or stand-alone data development, analysis, and production. In any of these paradigms, complete GIS functionality is available to a decision-maker, planner, resource manager, or other user.

This research effort is being conducted at the Center for Environmental Design Research under the sponsorship of the Department of Landscape Architecture and the State of California Department of Water Resources, with additional cooperation of the US Army Corps of Engineers, Construction Engineering Research Laboratory.

INTRODUCTION

The exploding interest in Geographic Information Systems that we are witnessing as we approach the 1990s is manifesting itself in users' demands for direct access to timely, accurate, and understandable geographic information. More and more, organizations are requiring that systems they acquire or develop make such information readily available to planners, decision-makers, and other users, who often are scattered throughout a region or state. An increasing emphasis on standards for communication and data exchange and on "user-friendliness" to a broad community of users is forcing systems developers to rethink traditional approaches to the design of GIS, and indeed to the entire range of computing resources.

The *Geographic Resource Analysis Support System* (GRASS) has been designed as a

high-performance, interactive environment for access to geographic data management, analysis, and display toolsGoran, 1987. Developed and distributed at cost by the Army Corps of Engineers, Construction Engineering Research Laboratory, in cooperation with several federal agencies and universities, GRASS utilizes the new generation of based graphics workstations, which provide mini and supermini computer performance in relatively low-cost desktop systems. This configuration provides organizations with complete GIS capabilities for both complex analytical processing and end-user access to map-based information.

X-Windows is a software system for computer graphics, windowing, and user interaction that has been developed over the last three years at Massachusetts Institute of Technology and placed into the public domain. X-Windows is unique in its strong network orientation and device-independence. This provides application developers and users with tremendous flexibility for transparently accessing software and data resources at remote locations on a heterogeneous network of computers.

The complimentary nature of GRASS's flexibility and X-Windows' network extensibility suggests the value of a merge of the two. By making the interactive capabilities of GRASS available for use throughout a fully distributed environment via the mechanism of X-Windows, some of the goals of direct user access to geographic information can begin to be realized. This merge, currently underway at UC Berkeley's Center for Environmental Design Research, is being performed in the context of the Kern Water Bank project of the California Department of Water Resources. The goal of this particular effort is to make a variety of thematic layers available for analysis and display at various DWR offices, capitalizing on the rapid expansion of the State's high-speed data communications network.

SYSTEM DESIGN CONSIDERATIONS

An Overview of GRASS

GRASS is a grid-based GIS for maintenance, analysis, and display of automated map information. It includes facilities for digitizing maps (in arc/node format), for importing existing vector and raster data, for performing boolean overlay, weighted modeling, tabulation and statistics, and other analyses, for displaying and interacting with graphic map information at a workstation, and for plotting and printing cartographic products. It also includes a subsystem for imagery analysis and integration of image and map data. A complete description of GRASS capabilities is beyond the scope of this paper, but a number of additional publications describe it in great detailCERL, 1987aCERL, 1987b. Although GRASS development is centered at CERL (the Army Corps' Construction Engineering Research Laboratory), the fact that GRASS was originally placed into the public domain has meant that other sites have made numerous extensions to the original system (see *GRASSClippings*, the quarterly newsletter published by CERL).

Since the focus of this paper is the integration of GRASS and X-Windows, it is important to understand the graphics environment in which GRASS operates from both a user's and a programmer's perspective. For a user, various display utilities and tools have been designed to provide a high level of user interaction. The user can display any cell map on the screen (or overlay multiple cell maps), overplot any number of vector or DLG files, and manipulate the resulting image with a combination of mouse, keyboard, and program-driven commands. For example, he or she can change any or all the colors displayed, identify real-world coordinates for locations on the screen, query the attributes associated with a point or region in any thematic layer in the database, zoom and pan, perform linear and areal measurements, generate label files, and correct or update a data layer. In addition, within the imagery subsystem, he or she can georeference imagery with other map data, select training sites, and perform various other display enhancement functions. A fundamental design goal of the GRASS-X-Windows merge project is to at a minimum maintain all existing capabilities at their existing level of performance, and to enhance or extend them wherever possible.

From a programmer's perspective, GRASS code has been extensively modularized in order to minimize the amount of

hardware dependent software. Written in C for the UNIX[†] environment, most of GRASS is essentially portable to a wide range of hardware. The use of only a few device-dependent routines thus means that the entire system can be ported without major effort to new platforms.[1] Isolation of hardware-dependent code is accomplished through the adoption of a *process-driver* modelWestervelt et al., 1987. All GRASS utilities that directly or indirectly relate to the display are written using general-purpose display subroutines. These subroutines are in turn represented in a separate library (*rasterlib* and *displaylib*, primarily), which characterizes the generic display primitives necessary to perform the desired operation. Finally, a separate library defines the specific system software hardware calls that will produce the graphic on the workstation screen. In existing configurations, these software/hardware calls are kernel-based, such as SunCore and Masscomp GKS.

At run-time, a background process—the GRASS *driver*—embodying the translation of the generic display primitives into the hardware-dependent system calls is started on the display device or within a display window. Display processes communicate directly with this driver via an interprocess communication (IPC) channel, generally *fifo*'s (System V UNIX so-called named pipes) or *sockets* (Berkeley UNIX networking). The rationale for this seeming complexity is that display tools are identical from machine to machine, and even on a specific computer, they may connect at run-time to various GRASS drivers (if for example a host was configured with multiple display monitors). As with the user environment, a primary goal of the GRASS X-Windows merge is to maintain the efficiency of the existing model.

The X-Windows Programming Environment

X-Windows is a graphics and windowing system for high-resolution workstations. It provides a user environment for creating and manipulating windows, using a keyboard and a mouse to provide input to applications software, and accessing multiple programs and systems concurrently. Because of numerous advantages to the system developer and its fundamental device independence, *X*-Windows is emerging as the standard for window systemsHack, 1987. From a programmer's perspective, *X*-Windows is a layered system allowing applications to be constructed on top of library subroutines and toolkit widgets, without regard for the underlying system operations.

At its most basic level, the *X*-Windows environment is analogous to that of GRASS. That is, utilities wishing to display text or graphics on a workstation screen communicate those requests to a separate process that is configured for a particular hardware platform. In the lexicon of *X*, these utilities are called *clients*, and the hardware dependent process is called the *server*. Client applications may exist on the same actual CPU as the *X* server, or they may be running on some other node on the network, that may not even have the same hardware architecture. *X*-Windows differs from other windowing and graphics systems precisely because of this reliance on a network protocolScheiflerGettys, 1986. All client-server communication is via simple, asynchronous byte streams sent over standard network protocols such as TCP/IP (*Transport Control Protocol/Internet Protocol*).

Because of this, all *X* clients are device independent, sending instructions (defined in *Xlib*) via network or interprocess communication to the *X* server on a local or remote host specified by a unique name or address. Servers run as continuous background processes on any addressable host, listening for display requests and mediating keyboard and mouse input. The servers then are the only display dependent part of *X*-Windows. For a user to use any *X*-based application, anywhere on a network, he or she needs only to ensure that the local display (ie, the workstation he or she is using) has a *X* server available and running.[2]

[†] UNIX is a trademark of Bell Laboratories

[1] Currently, GRASS runs on Sun, Masscomp, and AT&T 3B2 computers, using the native graphics system for each. Additional ports to Compaq, Macintosh, and Apollo are either planned or underway.

[2] This is not an academic or theoretical discussion. At this writing, workstations for which *X* is

Typical client applications include virtual terminal emulators, Tektronix graphics emulators, calendar and mail handling tools, text editors, and so on. A number of clients are designed to provide window/menu/button interfaces to existing general and special purpose applications packages.

A special case of client application is the user's *window manager*. By explicit design decision, *X* does not enforce any policy about what the *look-and-feel* should be. Instead, the window manager defines how a user manipulates windows, and may provide tools such as pop-up menus, scrollbars, titles and borders, and icons. The window manager must provide a mechanism for moving and cycling through a hierarchy of windows, and in conjunction with the *X* server itself deal with requests from applications for new windows, selecting graphics and text for pasting from one window to another, and refreshing the screen after a change. Of course, applications may have their own user interface rules that differ from each other and those of the window manager itself. This remains one of the thornier implementation issues for the *X* developer.[3]

Toolkits are collections of routines written using the native *Xlib* subroutines allow programmers to easily include such features (or "widgets") as scrollbars, menus, titlebars and borders into their applications. To an extent, these serve to create a consistent look-and-feel across multiple applications, since at least these widgets look and operate in the same way. They may also provide more special-purpose capabilities to the programmer, such as color table handling, image and bitmap manipulation, and event queue polling. At this time a number of toolkits are available, including the *X Toolkit* from MIT, Carnegie-Mellon's *Andrew*OguraNeuwirth, 1986, and others Lee, 1988.

available and commonly used include Digital Equipment MicroVax, Sun Microsystems, Hewlett-Packard, Apollo, IBM RT, and others; development is under way on a variety of microcomputers as well. Virtually all workstation vendors have committed to *X*-Windows in some form.

[3] There are now some attempts to define a user interface policy, including Sun/AT&T's *Open Look*AT&T (American Telephone & Telegraph Company), 1988, Hewlett-Packard's *New Wave*, and others.

<u>Integrating GRASS and X-Windows</u>

The actual programming effort involved in seemlessly integrating GRASS and *X*-Windows is necessarily a phased project. Over time, *X* functionality will be extended from simple display in a networked environment to completely new tools based on an *X*-Windows user interface. At the same time, the existing base of GRASS databases and users also requires that all modifications be fully upward compatible with current implementations. Moreover, GRASS procedures and display manipulation must continue to be available on platforms other than *X* workstations, for example ASCII *terminals* associated with specialized image subsystems. This implies the creation of a parallel, rather than replacement, display and interface system.

The initial phase, which is largely complete at this writing, is the development of a new display driver that communicates with an *X* server on a host. This driver simply takes the place of those that provide GRASS display using hardware-dependent graphics software. At run-time, the grass-*X* driver translates the graphics requests, communicated to it via fifo's from display utilities, into *X*-protocol syntax, and sends them to the designated server via sockets or IPC. In this initial phase, there is no change to the *rasterlib* which generates graphics requests from display subroutines. To a user, there is essentially no change from the existing environment-- one window acts as a control terminal and another acts as a display surface.

The next phase will involve a more complete integration of the two display mechanisms. This essentially requires a new version of *rasterlib* that is written using direct references to *Xlib* graphics calls, rather than to the device-independent subroutines currently used. By eliminating the "middleman" GRASS driver, the rewritten library should substantially increase graphics performance over the initial model. Moreover, the capabilities of this new "*Xrasterlib*" can be tuned to the specific features of the *X* server and rely on resources built into the server, such as fonts (for rapid labeling) or

backing stores of bitmaps (for interactive real-time graphics overlay).

As preliminary work has progressed on designing the new library, it has become clear that the its functionality takes on the characteristics of an *X* toolkit. As a special purpose toolkit, it would provide the GRASS programmer the building blocks necessary to implement display utilities, much as the existing *rasterlib* does now. In combination with general purpose toolkits, *Xrasterlib* would facilitate entirely new ways of implementing GRASS capabilities, including multiple display windows, more sophisticated use of pop-up menus, property sheets, and so on.

New user interface procedures will be implemented in a subsequent development phase, once the *Xrasterlib* has been proven robust and reliable. Gradually, as GRASS-X becomes more widely used and available, users at various sites will be able to generate their own *X*-based utilities and tools, relying on both general and special-purpose toolkits.

Of critical concern in any development program is the design of a user interface system that is flexible, powerful, and intuitive. For GRASS, also important is maintaining a consistent interface across the overall user environment and particular applications. In tandem, these restrictions may suggest the eventual implementation of a complete collection of toolkits that not only handle basic display requirements but also provide various widgets for the GRASS developer. Obviously the design and implementation of user interface toolkits will require a substantial commitment on the part of all GRASS users to reach consensus as to the nature of the best interface model. To some extent, the overall user interface design may be guided by various proposals being considered by Open Software Foundation and other groups.

THE GRASS-X PARADIGM

Models of X-based GIS

The use of a networked windowing system, such as *X*-Windows, in conjunction with any interactive GIS package, provides the application developer and user with several options for implementing a system. Such flexibility is ensured by *X*-Windows' client-server model, in which every system functions as a server with respect to certain capabilities and a client with respect to others. Thus a designated host can be a GRASS (or other GIS) applications server, a database server, a cartographic display server, or a server for any other desired purpose.

The first model of the GRASS-*X* implementation made possible by the work described in the previous section is remote operation with local X service. In this model, all geoprocessing and database analysis is performed on a host other than the users' workstations; the workstation handles only the actual on-screen display and mouse/keyboard input. The remote system supports both the application software and the spatial/attribute data for the GIS. The system may be optimized, in hardware and software, to perform GIS analysis more efficiently. Software needs to be supported (and possibly paid for) only at one site. The hardware may be of a completely different architecture, using a different operating system, than the *X*-server workstation. Indeed the local display system may be one of the new generation of *X* terminals, that provide network-speed[4] bit-mapped displays for *X*-client applications, but with no other local processing, for costs that have now dropped to below $1000 per system.

Ironically, this model is analogous to the older mainframe computer-graphics terminal systems that workstations are beginning to replace. In general, requiring one computer to perform all geoprocessing functions is less efficient than distributing that processing, but this may be a valid model in some circumstances. (For example, a broad community of users may make occasional use of GIS tools for simple queries or analyses, and even in their aggregate not overload a single GIS server.) Despite possible inefficiencies, this model is still superior in many respects to the mainframe-terminal analog, since the local workstations can still provide windows into other servers for other

[4] For example, Ethernet™ communication is at a maximum rate of 10 megabits/second, with typical throughput of 1 to 3 megabits/second.

applications, the communication with other systems is typically much faster than serial terminals, and more display processing is performed locally, offloading low-level graphics operations from the primary GIS host.

The second GRASS-*X* model comprises a distributed GIS implementation. It combines centralized GIS maintenance with local geoprocessing operations. A user at a workstation transparently accesses data residing on a remote host, performing any analytical or display operations on the workstation itself. Although invisible to the user, the needed data are actually transferred at high speed across the network to the workstation where they are manipulated by software that resides locally (or possibly on yet another host and transferred to the workstation). There are at least two technical methods of implementing the transfer of needed data and software: using relatively low-level operating system procedures, such as the industry-standard Network File System (NFS); or using intelligent data windowing or extraction techniques on the remote host. The first method is currently being used successfully for teaching and research projects at CEDR. It requires the least effort to develop and implement, but assumes that the remote data is directly manipulable within the local software environment.

For more complex situations, the second method should ultimately prove the better solution. User-generated requests for data would automatically be redirected to the remote host's native data retrieval subsystem. Appropriate spatial/attribute information would be extracted on the remote GIS server and effectively downloaded for local analysis and display. An extension to this method could also facilitate local database updates for uploading to the centralized server. As an example of this approach, a fully-supported GIS engine using *ARC/Info* for database development, maintenance, and cartographic production could provide query-determined datafile subsets in DLG format to a GRASS-*X* workstation located in a field office. There the user could perform rapid, interactive analyses and displays on the data without any further use of the *ARC/Info* server. Obviously, the systematic application of such technology is some time off, but there is a growing recognition of the need for sophisticated means for GIS data and software sharing.

The third model for using GRASS-*X* is a stand-alone mode. In this case, the *X*-based workstation provides all data development, analysis, display, and production functionality. No external data, software, or hardware is required. This model does not offer the advantages of using remote systems, but the local GIS user has complete control over the database once in place. A stand-alone system is probably most applicable for smaller jurisdictions or organizations will well-defined missions for their geographic information. With at least minimal communications capability, such sites could still obtain datasets from remote locations on an as-needed basis.

The three models presented are conceptual ones: in fact, any real-world application would almost certainly use some combination of the three. Some database components would be maintained locally, while others would be downloaded from centralized systems, and still others would be manipulated on an *ad hoc* basis on a range of distributed systems. The real value of the GRASS-*X* paradigm that has been outlined is in the ability to combine all three models in creative ways driven by individual user or organization requirements.

GRASS-X Advantages and Disadvantages

Although many aspects of the distributed models described above could be implemented under a variety of GIS and network systems, the specific features of both GRASS and *X*-Windows offer unique advantages to the system developer and user. In some situations these may be sufficient to provide complete GIS capability; in others they may be used as a gateway into even greater GIS functionality and data. Again, actual implementations will rely on a combination of these two modes.

GRASS has been designed from the outset to work well in a workstation environment, and thus is a logical vehicle for entering a distributed GIS. Especially when a number of installations are contemplated, GRASS has the additional advantage of being essentially free. This does imply

that organizations making a significant commitment to GRASS must also maintain the internal resources for installing and operating the software. Since GRASS is a very developmental system, new capabilities are added regularly and new subsystems must be supported.

Compared to other systems, GRASS is relatively easy to learn and use, and therefore can be used at remote sites without extensive internal support. Note that this observation applies to the use of GRASS *software*. It is critical to remember that intelligent use of GIS requires sophisticated understanding of the relationships among environmental features, of the nature of cartographic representations of the landscape, and of the explicit characteristics of the spatial/attribute database being used. As Nancy Tosta of the California Division of Forestry has observed (personal communication), "GIS does not support naive users." Easy to use software such as GRASS sometimes has the undesired side affect of enabling inappropriate use of complex GIS applications.

X-Windows is analogous to GRASS both in terms of its workstation orientation and its public domain status. It is available on virtually all major workstation platforms, and is beginning to appear on smaller systems as well. Since it is based on UNIX and on approved communications standards, it also lends itself to direct incorporation with GRASS.

Since *X*-Windows provides both network and windowing capabilities, it enables GRASS-*X* to operate in several modes simultaneously. That is, one display window may be showing a map generated from a local GRASS datafile, while another represents data on a remote system. Moreover, a variety of remote network nodes may be accessed at once through different windows, and in fact, a typical user session may easily utilize resources held on ten or more computers. Of course, these do not need to be only GIS applications or data— many types of tools, such as document processing, electronic mail, statistical analysis, and so forth, may be accessed concurrently. The local *X*-server handles the text, graphics, keyboard input, mouse manipulation, and other events automatically.

X is still a developmental system, and has not stabilized to the point of some commercial products. Although there is a commitment to upward compatibility, the very nature of its extensibility implies that somewhat divergent *X* environments will emerge. At this writing a number of developers are competing to establish common toolkits, user interface designs, and programming standards, so the future is not entirely clear. Vendors are also attempting to optimize *X*-servers, in particular, to their hardware environment, and sometimes hardware-independence suffers as a result.

Future Research Directions

Some of the major avenues for research are suggested in the preceding paragraphs, including an examination of the fundamental requirements for a flexible and powerful user interface and mechanisms for facilitating intelligent retrieval of information from diverse data sources around a network. Research is proceeding at Berkeley and elsewhere in numerous related domains, but little effort has been expended to date on relating these to GIS in general or to GRASS or *X*-Windows in particular. Fields of endeavor we hope to turn to in the future include:

Coordinate-Based Query. Work is already underway to allow use of digital maps as indices to a broad variety of information, including imagery, landscape photographs, archival materials, published and automated maps, etc. A logical extension of this capability is to enable coordinate-based query (ie., by pointing at a map location) of GIS data layers that may exist at any defined location within a network. Currently, a user must have prior knowledge of what data are available and where. Linking in an automated index could increase by orders of magnitude the volume and range of information effectively available at a site.

Integration with Hypermdia. Hypertext and hypermedia tools allow intuitive linkages to be built between disparate data elements; each element can be in a completely different form, including maps, images, text, spreadsheets, and so forth. Integrating these tools with fundamental GIS capabilities provides policy and decision makers with significantly more power to browse through

geographic information and explore relationships between environmental features.

Expert Systems. As the range of GIS data becomes more distributed and as GIS software becomes easier to use, more sophisticated *user agents* will be required. These user agents intelligently process queries, identifying the GIS data most appropriate to answering the question at hand and performing the necessary analyses. Expert systems technology is already in use in other domains, but the transfer of that technology to GIS is complicated by the very complexity of geographic information.

Object-Oriented Databases. Current research in database management systems (DBMS) is focussing on object-oriented databases. Object-orientation essentially defines every entity in the database as an object, encompassing features such as dimension, relation to other objects, hierarchical ascendancy and descendancy, topical connections, and so forth. Since geographic entities almost by definition include all these characteristics and more, they are a logical candidate for incorporation with object-oriented DBMS. Currently, work is underway on linking *postgres*, and object-oriented successor to the *ingres* relational DBMS, to geographic information.

Though these future directions are not necessarily tied to either GRASS or *X*-Windows, both systems appear to be appropriate vehicles for exploration. GRASS source code is freely available to researchers and can be modified without restriction; its modularity also lends itself to advanced work. *X*-Windows is also freely available, and much of the work in the areas outlined above is already based on *X*. By providing a strong linkage between GRASS and *X*, we hope to create a platform for examining these and other options.

CONCLUSIONS

The open systems approach of GRASS to GIS power, flexibility, and interactiveness is well-matched to the network extensibility and hardware-independence of *X*-Windows. This combination provides the GIS developer and user with the ability to perform complex geographic analyses on any node on a data communications network while displaying the results on a local workstation, to systematically extract information from distributed databases and analyze and display geographic information locally, and to maintain and manipulate entire GIS applications in a stand-alone environment. The combination also provides a useful platform for research into new GIS technologies. Together, the models for GRASS-*X* implementation and research constitute a new paradigm for distributed Geographic Information Systems.

REFERENCES

AT&T (American Telephone & Telegraph Company), 1988 AT&T (American Telephone & Telegraph Company) Open Look PM Graphical User Interface Functional Specifications July 1988 Sun Microsystems, Inc. 200 pages

CERL, 1987 CERL An Introduction to GRASS 13 pages. US Army Construction Engineering Research Laboratory Champaign, Ill 1987 Environmental Division

CERL, 1987 CERL Library of Resources GRASS/GIS US Army Construction Engineering Research Laboratory Champaign, Ill 1987 Environmental Division

Goran,1987 Goran GRASS and GIS Technology Corps Applications 1987

Hack,1987 Hack The Advantages of X Computer Graphics World 6, 57-60 August, 1987

Lee,1988 Lee Window of Opportunity UNIX Review 46-61 pages 1988

Ogura Neuwirth,1986 Ogura Neuwirth The Andrew System, Programmer's Guide to the Base Environment, Version 2 Information Technology Center, Carnegie Mellon University Pittsburgh 1986 IBM Corporation

Scheifler Gettys,1986 Scheifler Gettys The X Window System ACM Transactions on Graphics 79-109 1986

Westervelt et al.,1987 Westervelt Shapiro Higgins Larson Programmer's Refer-

ence Manual US Army Construction Engineering Research Laboratory Champaign, Ill 1987 Environmental Division

USACERL DISTRIBUTION

Chief of Engineers
ATTN: CEIM-SL (2)
ATTN: CECC-P
ATTN: CECW
ATTN: CECW-O
ATTN: CECW-P
ATTN: CECW-RR
ATTN: CEEC
ATTN: CEEC-C
ATTN: CEEC-E
ATTN: CERD
ATTN: CERD-C
ATTN: CERD-M
ATTN: CERM
ATTN: DAEN-ZCE
ATTN: DAEN-ZCI
ATTN: DAEN-ZCM
ATTN: DAEN-ZCZ

CEHSC ATTN: CEHSC-ZC 22060
ATTN: DET III 79906
ATTN: CEHSC-F 22060
ATTN: CEHSC-TF 22060
ATTN: Canadian Liaison Officer 65473
ATTN: German Liaison Staff 65473
ATTN: French Liaison Officer 65473
ATTN: Water Resources Center 22060

US Army Engineer Districts
ATTN: Library (41)

US Army Engr Divisions
ATTN: Library (14)

US Army Europe
ODCS/Engineer 09403
ATTN: AEAEN-FE
ATTN: AEAEN
V Corps
ATTN: DEH (11)
VII Corps
ATTN: DEH (16)
21st Support Command
ATTN: DEH (12)
USA Berlin
ATTN: DEH (9)
Allied Command Europe (ACE)
ATTN: ACSGEB 09011
ATTN: SHIHB/Engineer 09055
ATTN: AEUES 09081

USASETAF
ATTN: AESE-EN D 09019

8th USA, Korea (19)

ROK/US Combined Forces Command 96301
ATTN: EUSA-HHC-CFC/Engr

USA Japan (USARJ)
ATTN: DCSEN 96343
ATTN: Facilities Engineer 96343
ATTN: DEH Okinawa 96331

Area Engineer, AEDC-Area Office
Arnold Air Force Station, TN 37389

416th Engineer Command 60623
ATTN: Facilities Engineer

US Military Academy 10966
ATTN: Facilities Engineer
ATTN: Dept of Geography & Computer Science
ATTN: MAEN-A

AMC Dir., Inst., & Svcs
ATTN: DEH (23)

DLA ATTN: DLA-WI 22304

DNA ATTN: NADS 20305

FORSCOM
FORSCOM Engineer, ATTN: Spt Det.
ATTN: Facilities Engineer (27)

HSC
Ft. Sam Houston AMC 78234
ATTN: HSLO-F
Fitzsimons AMC 80045
ATTN: HSHG-DEH
Walter Reed AMC 20307
ATTN: Facilities Engineer

INSCOM - Ch, Instl. Div.
Arlington Hall Station (4) 22212
ATTN: Facilities Engineer
Vint Hill Farms Station 22186
ATTN: IAV-DEH

USA AMCCOM 61299
ATTN: AMSMC-RI
ATTN: AMSMC-IS

Military Dist of Washington
ATTN: DEH
Cameron Station (3) 22314
Fort Lesley J. McNair 20319
Fort Myer 22211

Military Traffic Mgmt Command
Falls Church 20315
Oakland Army Base 94626
Bayonne 07002
Sunny Point MOT 28461

NARADCOM, ATTN: DRDNA-F 01760

TARCOM, Fac. Div. 48090

TRADOC
HQ, TRADOC, ATTN: ATEN-DEH 23651
ATTN: DEH (18)

TSARCOM, ATTN: STSAS-F 63120

USAIS
Fort Huachuca 85613
ATTN: Facilities Engineer (3)
Fort Ritchie 21719

WESTCOM
Fort Shafter 96858
ATTN: DEH
ATTN: APEN-A

SHAPE 09055
ATTN: Survivability Sect. CCB-OPS
ATTN: Infrastructure Branch, LANDA

HQ USEUCOM 09128
ATTN: ECJ 4/7-LOE

Fort Belvoir, VA 22060
ATTN: British Liaison Officer
ATTN: Australian Liaison Officer
ATTN: Engr Studies Center
ATTN: Engr Topographic Lab
ATTN: ATZA-TE-SW
ATTN: STRBE-BLURE

CECRL ATTN: Library 03755

WES, ATTN: Library 3918

HQ, XVIII Airborne Corps and
Ft. Bragg 28307
ATTN: AFZA-DEH-EE

Chanute AFB, IL 61868
3345 CES/DE, Stop 27

Norton AFB, CA 92409
ATTN: AFRCE-MX/DE

Tyndall AFB, IL 32403
AFESC/Engineering & Service Lab

NAVFAC
ATTN: Division Offices (11)
ATTN: Facilities Engr Cmd (9)
ATTN: Naval Public Works Ctr (9)
ATTN: Naval Civil Engr Lab (3)
ATTN: Naval Constr Battalion Ctr

Engineering Societies Library
New York, NY 10017

National Guard Bureau 20310
Installation Division

US Government Printing Office 22304
Receiving/Depository Section (2)

US Army Env. Hygiene Agency
ATTN: HSHB-ME 21010

Nat'l Institute of Standards & Tech 20899

Defense Technical Info. Center 22314
ATTN: DDA (2)

316
09/89